Ottobre 1988
Numero 242
Anno XXI
Volume XLI

Il problema delle piogge acide

Recenti sviluppi tecnologici prospettano soluzioni vantaggiose dal punto di vista sia ambientale sia economico per contrastare efficacemente gli effetti deleteri delle deposizioni acide sul terreno e sulle acque

di Volker A. Mohnen

L'atmosfera funge da camera di reazione chimica per una grande varietà di sostanze. Molte delle più importanti, come ossigeno, anidride carbonica e composti dell'azoto e dello zolfo, vengono prodotte dall'attività degli organismi; attraverso il ciclo dell'acqua le sostanze liberate nell'atmosfera ritornano alla fine nel terreno, nelle acque superficiali e nella materia organica. Lo sviluppo industriale ha accresciuto enormemente il «carico» nell'atmosfera di alcune di queste sostanze, con conseguenze di vasta portata per la vita e l'ambiente. La dimostrazione più lampante si ha nel caso delle piogge acide, precipitazioni miste a particellato acidificante dall'inquinamento atmosferico.

L'allarme per la crescente acidità delle precipitazioni in Europa e in America Settentrionale fu lanciato per la prima volta all'inizio degli anni sessanta. Da allora gran parte dell'attenzione si è concentrata sugli effetti delle piogge acide, accertati o probabili, su laghi e corsi d'acqua, incluse le forme di vita da essi ospitate, e sulle foreste, anche se la lista delle situazioni preoccupanti è assai più lunga: comprende la contaminazione delle acque di falda, la corrosione di strutture costruite dall'uomo e, più di recente, il deterioramento delle acque marine lungo le coste. A 20 anni dai primi rilevamenti, in quale misura i danni all'ecosistema, in particolare ai laghi e alle foreste, sono stati accertati e misurati? Cosa si è appreso sui processi che producono le piogge acide e ne determinano gli effetti dannosi? Che tipo di sforzi implicano queste conoscenze per il controllo delle emissioni che danno luogo alle deposizioni acide, in particolare quelle di biossido di zolfo dalle centrali a carbone e a gasolio e quelle di ossidi di azoto dai motori dei veicoli e ancora dalle centrali termoelettriche?

Gli studi per rispondere a queste domande hanno assunto le dimensioni di una grande impresa scientifica. Sotto l'egida del National Acid Precipitation Assessment Program (NAPAP), reso esecutivo nel 1980, vari Enti governativi sovvenzionano ricerche sui processi atmosferici che danno origine alle piogge acide, sui loro effetti sull'ecosistema e sulle possibilità di controllare i processi stessi. Inoltre l'Electric Power Research Institute, che riceve fondi dai produttori di energia, promuove studi sugli effetti delle precipitazioni acide e su tecnologie atte a ridurre le emissioni delle centrali termoelettriche. Il NAPAP non pubblicherà un rapporto completo prima del 1990, ma sono già disponibili dati sufficienti per affermare che le piogge acide, o più correttamente gli inquinanti che le provocano, rappresentano un'interferenza su vasta scala nei cicli biogeochimici tramite i quali gli esseri viventi interagiscono con l'ambiente. Una corretta gestione globale richiede uno sforzo per proteggere l'integrità di questi cicli e ora abbiamo i mezzi economici per farlo.

Le precipitazioni acide sono una diretta conseguenza dei processi di autodepurazione dell'atmosfera. Le minuscole goccioline d'acqua che costituiscono le nubi «catturano» continuamente particelle in sospensione e gas solubili in traccia e l'acqua piovana che da esse si forma per coalescenza rimuove le impurezze dall'atmosfera. Non tutti i gas presenti in tracce possono essere eliminati dalle precipitazioni, ma il biossido di zolfo (SO_2) e gli ossidi di azoto subiscono una trasformazione chimica rispettivamente in acido solforico e nitrico, che possono essere facilmente incorporati nelle goccioline d'acqua.

I processi che trasformano gli ossidi gassosi in acidi e li rimuovono dall'atmosfera erano operanti molto prima che avesse inizio l'uso massiccio di combustibili fossili; composti dello zolfo e dell'azoto vengono prodotti anche da processi naturali quali il vulcanismo e l'attività dei batteri nel suolo. Ma le attività economiche umane hanno notevolmente accresciuto l'importanza di queste reazioni, innescate dalla luce solare e dipendenti dalla disponibilità di acqua e di ossigeno nell'atmosfera.

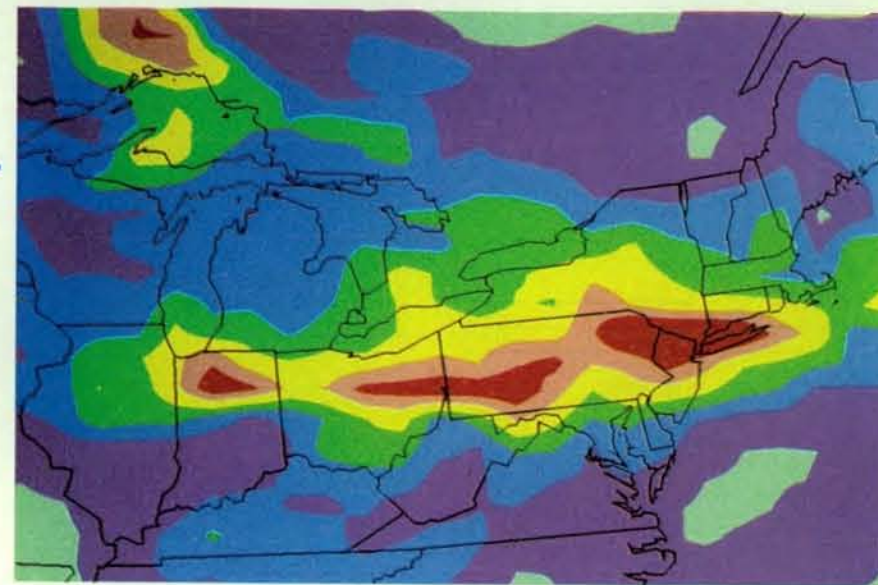
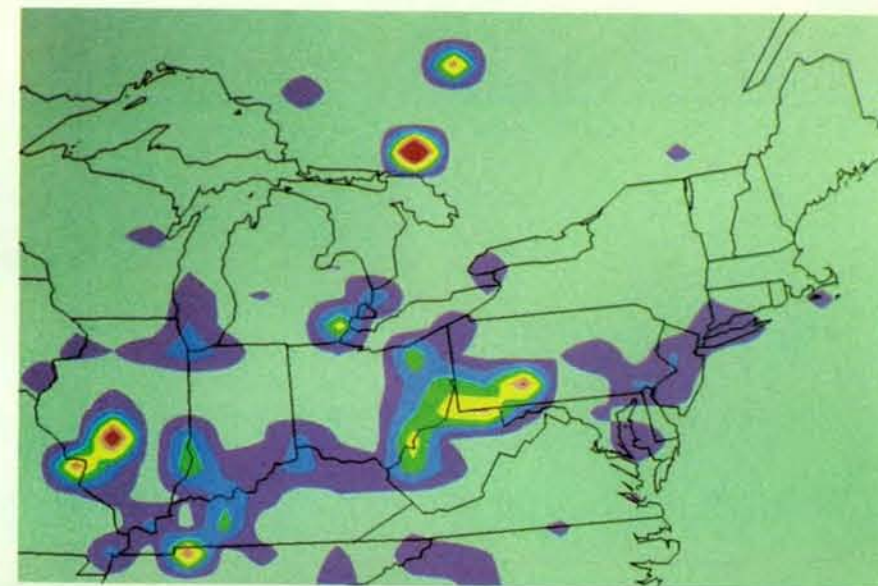
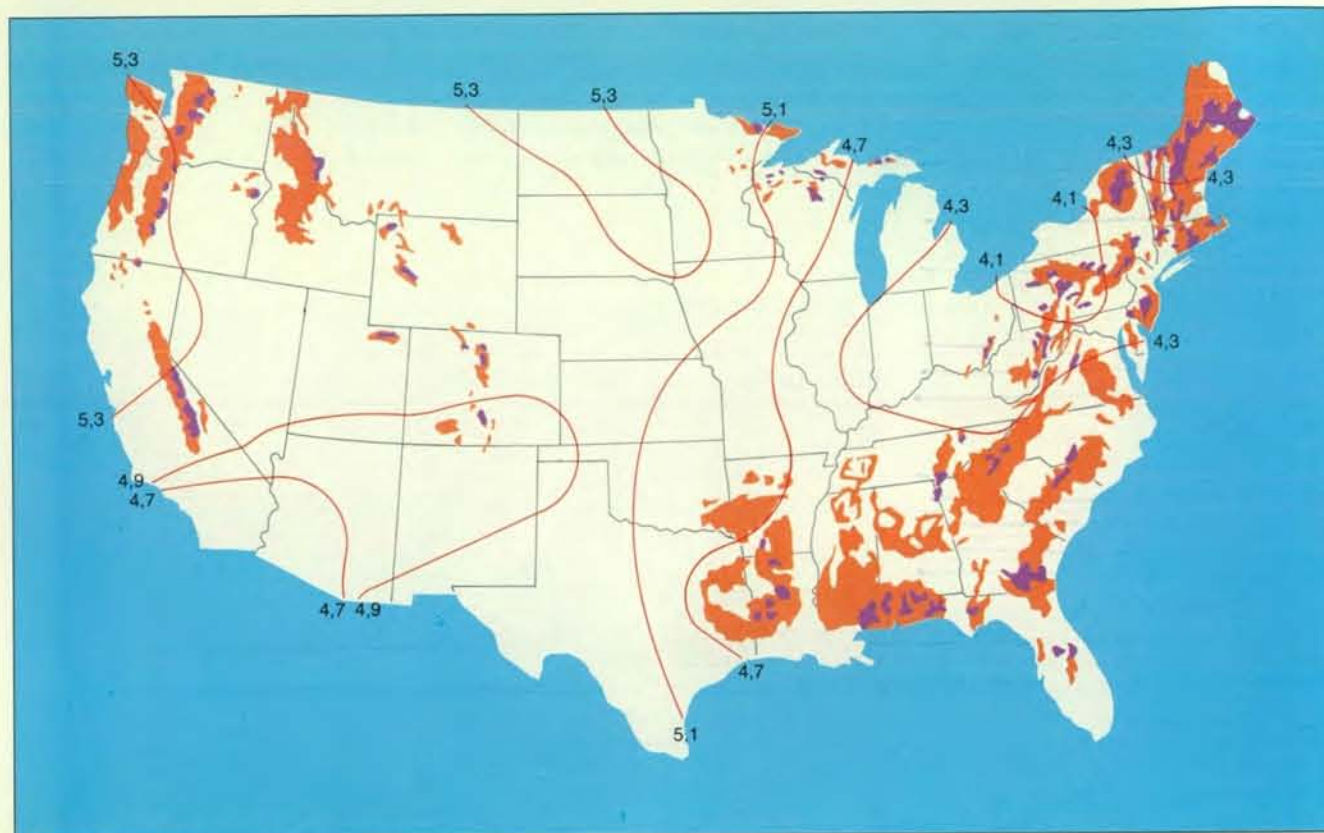
Il ciclo di reazioni avviene nella troposfera, la porzione di atmosfera che si estende dal suolo fino a un'altezza di 10-12 chilometri. Esso inizia quando un fotone di luce solare colpisce una molecola di ozono (O_3), che può provenire dallo strato di ozono presente nella stratosfera o può essersi formata nella troposfera stessa per azione di inquinanti contenenti azoto o carbonio. I prodotti della reazione sono una molecola di ossigeno (O_2) e un atomo di ossigeno isolato, altamente reattivo, che si combina con una molecola d'acqua (H_2O) a formare due radicali ossidrile ($OH\cdot$). Questa specie chimica, poco abbondante ma reattiva, trasforma il biossido d'azoto (NO_2) in acido nitrico (HNO_3) e funge da iniziatrice delle reazioni che trasformano il biossido di zolfo in acido solforico (H_2SO_4).

La concentrazione del radicale ossidrile nell'atmosfera è inferiore a una parte per bilione, ma è praticamente costante, dato che parecchi dei processi di ossidazione che esso promuove finiscono per rigenerarlo. Per esempio, un sottoprodotto dell'ossidazione iniziale del biossido di zolfo è il radicale perossidrile ($HO_2\cdot$), che reagisce con il monossido d'azoto (NO) a dare biossido d'azoto e un nuovo radicale ossidrile. In effetti un solo radicale ossidrile può ossidare migliaia di molecole contenenti zolfo; ne deriva che la quantità di acido prodotta dipende solo dalla quantità di biossido di zolfo presente nell'aria.

L'acido solforico e l'acido nitrico che derivano dagli inquinanti gassosi possono essere facilmente incorporati nelle nubi. (Una certa quantità di acido solforico si forma direttamente nelle goccioline contenute nelle nubi, a partire dal biossido di zolfo disciolto e dal perossido d'idrogeno.) L'acido nitrico gassoso è facilmente solubile in goccioline già formate. L'acido solforico che si produce nel corso di reazioni in fase gassosa condensa a formare goccioline microscopiche, di diametro variabile fra 0,1 e 2 micrometri (milionesimi di metro), che sono uno dei componenti delle foschie estive negli Stati Uniti orientali. Alcune di queste particelle raggiungono il terreno nel processo definito di «deposizione a secco». (Con questo nome si indica anche l'assorbimento di biossido di zolfo gassoso da parte della vegetazione.) La maggior parte di esse, tuttavia, viene incorporata nelle nubi. L'umidità condensa facilmente su una superficie che funge da nucleo di condensazione e le particelle di solfato costituiscono nuclei ideali; le gocce di pioggia che si sviluppano intorno a esse contengono acido solforico diluito.

Gli acidi solforico e nitrico disciolti possono conferire alle gocce di pioggia un *p*H estremamente basso. I campioni di acqua raccolti presso la base delle nubi negli Stati Uniti orientali, durante l'estate, hanno generalmente *p*H pari a 3,6 circa, ma sono stati registrati valori anche di 2,6. (Il *p*H neutro è uguale a 7; valori via via più bassi corrispondono ad acidità crescenti.) Nell'area metropolitana di Los Angeles il *p*H della nebbia è sceso a 2, valore paragonabile a quello del succo di limone.

Acidità così elevate si riscontrano solo alla base delle nubi; le zone più elevate sono considerevolmente più «pulite». Quando il suolo e la vegetazione vengono avviluppati dalle nubi acide, come può accadere nelle foreste di alta montagna, si trovano esposti direttamente alla zona di elevata acidità alla base delle nubi. Le precipitazioni, invece, raccolgono acqua per tutta l'altezza della nube. La conseguente diluizione abbassa la concentrazione dei composti dello zolfo e dell'azoto di un fattore compreso fra tre e 30 e l'acidità diminuisce di un valore che va da mezza a una unità *p*H, fino a un *p*H medio 4,2 nella zona nordorientale degli Stati Uniti.

Le piogge acide possono cadere a centinaia di chilometri di distanza dalla fonte di inquinamento. Nella zona di caduta ha inizio una nuova serie di muta-

Deposizioni di acido solforico in tre giorni di pioggia rappresentate mediante un modello al calcolatore. L'immagine in alto mostra le emissioni di zolfo (soprattutto biossido di zolfo) negli Stati Uniti nordorientali e nel Canada meridionale in base alle quali sono stati introdotti i dati per elaborare il modello. Il colore verde chiaro corrisponde ai livelli di emissione più bassi e il rosso ai più elevati. Le centrali a carbone nel Middle West costituiscono la principale fonte di emissioni degli Stati Uniti, mentre la fonderia di rame-nichel a Sudbury nell'Ontario rappresenta la fonte più cospicua in territorio canadese. Basandosi sulle condizioni meteorologiche dei giorni fra il 22 e il 24 aprile 1981, il calcolatore ha determinato il trasporto dei composti dello zolfo e di altre sostanze chimiche correlate, la loro trasformazione in acido solforico e la loro deposizione nell'arco dei tre giorni. L'immagine in basso mostra come l'acido solforico si è depositato con la pioggia; il verde chiaro indica aree che hanno ricevuto meno di 10 grammi di zolfo per ettaro e il rosso indica zone che ne hanno ricevuto più di 260 grammi. Il modello al calcolatore, detto Regional Acid Deposition Model, è stato messo a punto da Julius Chang e colleghi della State University of New York ad Albany, con finanziamenti della US Environmental Protection Agency.

La carta mostra la minaccia rappresentata dalle piogge acide per laghi e corsi d'acqua negli Stati Uniti. Il colore arancione identifica aree in cui le acque superficiali tendono ad avere bassa alcalinità (ossia uno scarso contenuto di ioni che neutralizzano gli acidi, come il bicarbonato), il viola quelle in cui l'alcalinità è minima. Le linee in rosso indicano il *p*H medio delle precipitazioni. Dove le piogge acide interessano acque a bassa alcalinità, laghi e fiumi rischiano l'acidificazione (l'alcalinità non è tuttavia il solo fattore da cui dipende la sensibilità alle piogge acide). I dati sull'alcalinità sono stati forniti da James M. Omernik e colleghi dell'EPA.

menti fisici e chimici, che può ridurre l'acidità e modificare le caratteristiche chimiche dell'acqua che infine raggiunge laghi e torrenti. I suoli alcalini, come quelli contenenti rocce calcaree, possono neutralizzare direttamente l'acidità. Nei suoli leggermente acidi tipici delle foreste di conifere degli Stati Uniti, del Canada e dell'Europa, altri due processi possono moderare gli effetti delle deposizioni acide. L'acidità può essere «immobilizzata» quando il suolo o la vegetazione trattengono ioni solfato o nitrato (derivati dall'acido solforico e nitrico rispettivamente), oppure può essere «tamponata» tramite lo scambio di ioni positivi.

In quest'ultimo processo gli ioni calcio, magnesio e di altri metalli che si trovano in diversi suoli prendono il posto degli ioni idrogeno nell'acido. Gli ioni metallici vengono prodotti dalla disgregazione delle rocce, ossia la dissoluzione di minerali da parte delle acque piovane e sotterranee contenenti anidride carbonica disciolta, che liberano nel suolo gli ioni metallici positivi insieme ad anioni bicarbonato (HCO_3^-). Quando poi l'acido solforico giunge nel suolo, l'anione solfato (SO_4^{2-}) può spostare gli ioni calcio o magnesio. Mentre gli ioni metallici vengono rimossi dal suolo da parte della soluzione contenente solfato, gli ioni idrogeno, responsabili dell'acidità, rimangono *in loco*.

L'entità della ritenzione nel suolo e dello scambio di cationi nelle acque superficiali o sotterranee dipende dalle caratteristiche del bacino idrico, in particolare dalla geologia, dalla vegetazione e dall'andamento del flusso. I processi che avvengono nel suolo non possono avere effetti su acque che scorrono su terreni gelati o saturi o su un basamento di granito e perciò l'acqua che raggiunge un lago o un torrente in questi casi rimane acida quanto quella piovana. Anche quando l'acqua riesce a infiltrarsi in profondità i processi in questione possono risultare inefficaci. Il quarzo, per esempio, è resistente alla disgregazione ed è privo degli ioni metallici necessari per lo scambio di cationi; di conseguenza la percolazione attraverso sabbia quarzosa contribuisce ben poco a moderare l'acidità. Tuttavia, in un bacino idrico avente un suolo profondo in grado di trattenere grandi quantità di solfati e nitrati, o ricco di cationi facilmente scambiabili, il trasferimento di acidità in un lago o in un corso d'acqua può essere rimandato, almeno fino a quando le capacità di ritenzione o di effetto tampone non siano esaurite.

Che cosa accade quando acque superficiali o sotterranee acide raggiungono un lago o un torrente? Un corpo idrico può contenere bicarbonato e altri ioni alcalini derivati dalla disgregazione delle rocce, che neutralizzano le sostanze acide in ingresso e impediscono al *p*H dell'acqua di scendere al di sotto di 5. La cosiddetta capacità di neutralizzazione dell'acqua, ossia il suo contenuto in ioni in grado di neutralizzare l'acidità, rappresenta una misura della suscettibilità di un lago all'acidificazione. Un lago che presenta una capacità di neutralizzazione elevata è almeno momentaneamente protetto dagli effetti delle piogge acide; al contrario, un lago avente capacità di neutralizzazione nulla può rimanere «sano» solo se non è esposto a precipitazioni acide, altrimenti viene acidificato immediatamente.

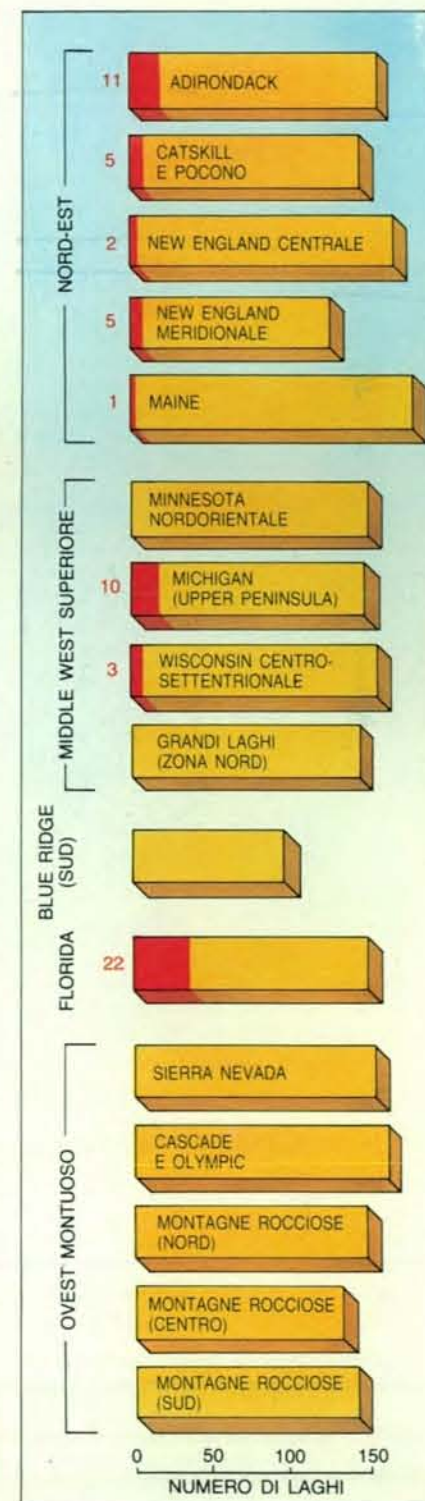
È facile individuare un lago acidificato. La sua capacità di neutralizzazione è ridotta a zero e il *p*H presenta un valore inferiore a 6; le acque contengono concentrazioni elevate di ione solfato e di altri ioni, come l'alluminio, che vengono mobilizzati dalla percolazione di soluzioni acide nel suolo. Questi laghi ospitano comunità biologiche anomale (o addirittura sono privi di forme di vita). Prevedere l'acidificazione di un lago che abbia una capacità di neutralizzazione bassa ma ancora rilevabile è invece un compito meno semplice. La ritenzione o l'effetto tampone sulle sostanze acide che si depositano nel bacino idrico possono per un certo tempo rallentare la diminuzione della capacità di neutralizzazione. Inoltre un lago non ha una quota fissa di ioni in grado di neutralizzare l'acidità: anche se viene depauperata dall'ingresso di sostanze acide, la capacità di neutralizzazione può ritornare a valori più elevati a causa della disgregazione delle rocce nei dintorni del lago. Per prevedere come un lago potrà reagire a un apporto costante di sostanze acide è necessario conoscere non solo la sua capacità di neutralizzazione, ma anche quanto velocemente essa può venire ripristinata e quanto a lungo il tasso di ripristino può essere mantenuto.

Questi processi che avvengono tra bacino idrico e corpo idrico determinano perciò se un dato lago subirà acidificazione e a quale tasso. Siamo ben lungi dal comprenderli perfettamente e cercare di conoscere un sistema abbastanza bene da prevederne il comportamento è difficile. Tuttavia, non ci sono dubbi sulla tendenza generale: nelle regioni in cui il suolo è povero di minerali disgregabili e le precipitazioni sono fortemente acide i laghi si stanno acidificando. Nel 1986 un comitato della National Academy of Sciences ha raccolto misure di *p*H e alcalinità (una misura del potere tampone simile alla capacità di neutralizzazione) eseguite tra il 1920 e il 1950 in diverse centinaia di laghi negli Stati del Wisconsin, del New Hampshire e di New York e le ha confrontate con dati recenti. Il comitato ha concluso che nel detto periodo il *p*H e l'alcalinità erano in media aumentati nei laghi del Wisconsin mentre erano rimasti generalmente invariati in quelli del New Hampshire. Nello Stato di New York, tuttavia, e in particolare nei monti Adirondack, i dati relativi ad alcuni laghi mostravano una netta tendenza verso l'acidificazione.

Il comitato ha ottenuto un quadro più completo dallo studio dei microrganismi conservati nei sedimenti sul fondo dei laghi. Quando il *p*H di un lago varia, si modifica anche la popolazione di diatomee e di alghe brune che esso ospita. Le diverse specie di diatomee possono essere distinte con l'esame dei loro scheletri silicei e questo fatto permette di ricostruire le variazioni nel tempo della comunità di specie e quindi del *p*H dell'acqua. Degli 11 laghi nei monti Adirondack per i quali erano disponibili dati, sei avevano iniziato ad acidificarsi negli anni trenta e avevano raggiunto un *p*H inferiore a 5,2; l'acidificazione più veloce si era avuta nel periodo precedente al 1970. L'unica causa di abbassamento del *p*H che il comitato fu in grado di identificare era appunto la pioggia acida.

L'acidificazione dei laghi nei monti Adirondack è provocata dalle precipitazioni altamente acide che cadono nella regione (le piogge campionate nei dintorni, nella parte occidentale dello Stato di New York, hanno un *p*H medio di circa 4,1, il più basso di tutto il paese) e dallo scarso potere tampone del suolo e dei laghi a basamento granitico. Nell'ambito del recente National Surface Water Survey sono state esaminate altre zone in tutto il paese dove la capacità di neutralizzazione di laghi e corsi d'acqua tende a essere bassa e li rende vulnerabili alle piogge acide. Il rilevamento ha permesso di individuare una notevole percentuale di laghi acidificati sui monti Pocono, nella Pennsylvania orientale, e nella Upper Peninsula del Michigan, due regioni in cui le precipitazioni sono notevolmente acide. Le piogge acide sono ritenute con tutta probabilità responsabili del numero relativamente elevato di laghi acidificati nel New England centrale e meridionale. In Florida vi è una percentuale sorprendentemente alta di laghi acidificati, ma si pensa che questo sia dovuto ad altre circostanze, come la produzione di acidi organici in seguito alla decomposizione della vegetazione nelle aree paludose e l'arricchimento in fertilizzanti delle acque che scorrono attraverso le zone agricole.

Il Maine ha la più bassa percentuale di laghi acidificati nella regione nordorientale, nonostante lo scarso potere tampone dei suoli e delle acque. Anche nell'alta regione dei Grandi Laghi, nelle Blue Ridge Mountains meridionali e nelle regioni montuose occidentali specchi d'acqua in analoghe condizioni sono risultati per lo più sani, con un *p*H superiore a 6. Ciò che distingue queste regioni è la relativa scarsità di precipitazioni acide.

Dati non altrettanto conclusivi sono stati raccolti finora per quanto riguarda l'altro importante effetto ambientale attribuito alle piogge acide: il declino delle foreste. Dal 1980 molte foreste negli Stati Uniti orientali e in varie parti dell'Europa hanno subito una drastica perdita di vitalità non collegabile ad alcuna delle cause normali, quali insetti, malattie o avvelenamento da parte di specifici inquinanti nell'aria e nell'acqua. La situazione più drammatica si riscontra in Germania, dove gli scienziati, stupiti per l'ampiezza e la velocità del declino, lo hanno definito *Waldsterben*,

I laghi acidificati degli Stati Uniti sono concentrati nella parte nordorientale e nel Middle West settentrionale. Il diagramma mostra i dati raccolti dal National Surface Water Survey. Il tratto in rosso in ciascuna barra rappresenta il numero di laghi studiati il cui contenuto di bicarbonato e di altri ioni in grado di neutralizzare gli acidi è ridotto; il loro numero è dato anche come percentuale (*in colore*). Questi laghi hanno di solito un *p*H basso e comunità biologiche anomale. Si ritiene che la causa principale dell'acidificazione siano gli acidi solforico e nitrico derivati dagli inquinanti. In Florida, invece, i fattori responsabili sarebbero gli acidi organici dovuti alla decomposizione della vegetazione e all'uso dei concimi.

o «morte delle foreste»; anche le statistiche che riguardano gli Stati Uniti sono comunque preoccupanti.

Il declino più evidente si ha nelle foreste di conifere ad alta quota. In parecchi siti posti a una quota superiore a 850 metri nei monti Adirondack, nelle Green Mountains del Vermont e nelle White Mountains del New Hampshire un confronto fra dati storici e attuali dimostra che negli ultimi 25 anni si è avuta la morte di più del 50 per cento degli abeti rossi. Ad altitudini minori sono stati documentati danni a specie di alberi dal legno sia tenero sia duro.

Almeno nelle foreste ad alta quota, i tronchi morti sono solo la manifestazione più evidente di una generale perdita di vitalità da parte degli alberi. Lo studio degli anelli di accrescimento annuali in foreste di montagna nella parte nordorientale degli Stati Uniti indica che, a partire dall'inizio degli anni sessanta, si è avuta una marcata riduzione nello spessore dello strato di accrescimento annuale. Il declino colpisce piantagioni da legname di varie età e con storie diverse di inconvenienti e malattie. Quale potrebbe essere il fattore comune responsabile della riduzione di crescita?

Il ruolo delle piogge acide e di altre forme di inquinamento atmosferico è oggetto di attive indagini; tuttavia, nonostante l'ampiezza dei danni alle foreste, un collegamento certo di causa ed effetto si è rivelato difficile da dimostrare. Si può avere un'idea delle difficoltà confrontando il recente declino delle foreste con casi accertati di avvelenamento da parte di fumi inquinanti. Fonderie e impianti chimici che emettono biossido di zolfo, ossidi di azoto o fluoruri sono spesso circondati da alberi morti. In questi casi vi è una netta correlazione fra il danno agli alberi, una fonte specifica di inquinamento e valori soglia nella concentrazione degli inquinanti. Le foreste che stanno morendo, invece, sono lontane da fonti di inquinamento e si trovano esposte a concentrazioni di sostanze nocive assai inferiori ai livelli che, in base ai dati raccolti, provocano danni diretti agli alberi. Se l'inquinamento atmosferico, e le precipitazioni acide in particolare, hanno un ruolo nel declino delle foreste, probabilmente essi agiscono più come fattori traumatici che come agenti letali.

Molti di questi fattori, di tipo biologico e non, possono combinarsi per minare la vitalità di una foresta. Un possibile fattore è dato dal corredo genetico o dall'età degli alberi: una piantagione può essere geneticamente debole o senescente. Altri traumi possono essere rappresentati da malattie, insetti, funghi e piante parassite, scarsità di luce, acqua o sostanze nutritive essenziali e danni occasionali dovuti a inondazioni, forti venti e tempeste di ghiaccio. Fattori di per sé facilmente sopportabili possono combinarsi tra loro con effetti debilitanti o letali. Una sequenza di eventi che conduce alla morte di un albero può iniziare con un trauma «predisponente», come una scarsità di sostanze nutritive. L'albero può quindi essere seriamente indebolito da un fattore «scatenante», come un inverno molto rigido, e trovarsi privo di difese contro l'eventuale trauma finale «concomitante» (la causa immediata della morte), quale una malattia o una infestazione da parte di insetti.

L'acidità e gli altri inquinanti potrebbero sommarsi all'alta incidenza di fattori traumatici non biologici, come suoli poco profondi, basse temperature e venti secchi (quindi disidratanti), che si incontrano nelle foreste di montagna. Questo significa che le sostanze nocive potrebbero caricare gli alberi del peso di un ulteriore fattore predisponente nel momento in cui essi devono fronteggiare successivi traumi. Ma qual è la natura di questi fattori aggiuntivi?

Parecchi studiosi, per lo più europei, hanno proposto diversi meccanismi ipotetici, molti dei quali darebbero come risultato una scarsità di apporti nutritivi per l'albero. Diversi meccanismi svolgerebbero la loro azione nel suolo. L'alluminio liberato dai minerali contenuti nel terreno per opera degli acidi potrebbe competere con il calcio per i siti di legame sui peli radicali, riducendo l'assorbimento di calcio da parte degli alberi e rallentandone la crescita. In alternativa, il suolo stesso potrebbe perdere sostanze nutritive per il dilavamento a opera delle piogge acide di elementi indispensabili alla vita quali calcio, magnesio e potassio. La morte dei microrganismi presenti nel suolo potrebbe essere un'ulteriore causa di riduzione dell'apporto nutritivo. Un pH basso e un'elevata concentrazione di alluminio nel suolo possono ridurre le popolazioni dei batteri che decompongono e liberano le sostanze nutritive contenute nei materiali organici in decomposizione. Per di più, alti livelli di nitrati, derivati dall'acido nitrico, possono danneggiare le micorrize, funghi simbionti che vivono sulle radici delle conifere e aiutano l'albero a combattere le malattie e ad assorbire acqua e sostanze nutritive.

Secondo altre ipotesi, le sostanze inquinanti avrebbero effetti soprattutto al di fuori del suolo. Le piogge acide o, più probabilmente, le goccioline acide nelle nubi intercettate dagli aghi di una conifera potrebbero lisciviarne le sostanze nutritive (in particolare magnesio, calcio e potassio) più velocemente di quanto possano essere rimpiazzate dalle radici. Un altro inquinante, l'ozono, è forse in grado di aggravare la perdita di sostanze nutritive degradando il rivestimento ceroso impermeabile degli aghi. Secondo un'ulteriore ipotesi, l'ozono da solo potrebbe provocare un trauma da scarsità di apporti nutritivi perché può danneggiare la clorofilla e impedire quindi la fotosintesi.

Infine, le piogge acide potrebbero aggravare i traumi dovuti alle basse temperature invernali. In autunno, di norma, una conifera si prepara al gelo invernale diminuendo il contenuto d'acqua negli aghi. Il segnale iniziatore viene dalle radici, sotto forma di una diminuzione delle sostanze nutritive azotate che sono prodotte dai microrganismi del suolo. Se una soluzione acida penetra negli aghi, tuttavia, i composti azotati che essa contiene potrebbero «concimare» l'albero. Se questo effetto sovrastasse il segnale proveniente dalle radici la preparazione all'inverno potrebbe essere ritardata, rendendo in tal modo l'albero vulnerabile ai danni dovuti alla formazione di ghiaccio nel tessuto degli aghi. Anche l'ozono può ridurre la resistenza di un albero al gelo, danneggiando le membrane delle cellule fogliari.

Sono attualmente in corso esperimenti in laboratorio per stabilire se qualcuno di questi meccanismi (e quale) possa essere operante nelle condizioni di esposizione agli inquinanti delle foreste colpite. Ma solo studi condotti sul campo, nelle foreste stesse, possono chiarire se un dato meccanismo sia effettivamente responsabile dei danni osservati. Il compito è soverchiante: bisogna cercare di riconoscere un incremento forse relativamente piccolo di fattori traumatici nell'ambito di tutta una serie di complessi fattori naturali, che potrebbero variare da piantagione a piantagione o anche da albero ad albero.

Il caso di Whiteface Mountain, nei monti Adirondack, è particolarmente eloquente. Il sito presenta uno degli esempi più drammatici di declino delle foreste negli Stati Uniti, ma a causa della preminenza di svariati fattori traumatici naturali si possono trarre solo conclusioni ipotetiche sul ruolo degli inquinanti. La causa diretta del declino della foresta, in base ai registri delle guardie forestali e ai dati di temperatura, sembra essere stata una serie di gravi e ripetuti danneggiamenti a causa dell'aridità o del gelo durante gli inverni dei primi anni sessanta. Anche l'ozono può avere reso gli alberi più vulnerabili alle basse temperature; le variazioni biochimiche che si verificano nella chioma di un albero quando si proteggono i rami dall'ozono atmosferico fanno pensare che questo gas possa in effetti indebolire l'albero attaccando le membrane delle cellule fogliari. Il ruolo delle precipitazioni e delle nubi acide non è stato ancora studiato a fondo, ma è probabile che anch'esse abbiano agito in qualche modo come fattori predisponenti.

Sebbene il ruolo esatto delle piogge acide nel declino delle foreste sia ancora oggetto di controversia, gli effetti di queste sul suolo e sulle acque non lasciano dubbi sulla necessità di ridurre il carico di composti dello zolfo e dell'azoto nell'ambiente e quindi di abbassare l'acidità delle precipitazioni. Alcuni progressi sono già stati compiuti. Nella regione nordorientale degli Stati Uniti il contenuto di solfati nelle piogge e la concentrazione di composti dello zolfo nell'atmosfera

Le reazioni chimiche che avvengono nell'atmosfera producono acido solforico e nitrico a partire da biossido di zolfo e ossidi di azoto emessi da industrie e autoveicoli. Il radicale ossidrile, che si forma quando una molecola di ozono si dissocia e libera un atomo di ossigeno, è il reagente principale. Esso trasforma il biossido di azoto (NO_2) in acido nitrico (*freccia blu*) e funge da iniziatore per la conversione del biossido di zolfo (SO_2) in acido solforico (*frecce rosse*). (In un'altra sequenza di reazioni l'acido solforico si forma dal biossido di zolfo e dal perossido di idrogeno in soluzione nelle nubi.) Il radicale ossidrile è rigenerato da reazioni (*frecce verdi*) a cui partecipa il monossido d'azoto (NO); gli acidi arrivano al suolo sia come particellato secco sia con la pioggia o altre precipitazioni.

I processi che avvengono nei bacini idrici possono modificare il chimismo dell'acqua piovana acida prima che questa raggiunga un lago o un corso d'acqua. L'illustrazione mostra i diversi processi che possono agire sull'acido solforico che percola lungo un versante. Se il suolo è ricco di cationi, per esempio calcio e magnesio (Ca^{2+} e Mg^{2+}), si può avere scambio di cationi. Gli ioni positivi sono liberati da certe rocce per l'azione disgregatrice dell'acqua sotterranea e piovana contenente anidride carbonica disciolta, un processo che genera anche ioni bicarbonato (HCO_3^-). Alcuni degli ioni idrogeno dell'acido (*in rosso*) spostano il calcio e il magnesio e sono a loro volta trattenuti nel suolo, dove gli ioni bicarbonato li neutralizzano. Questi fenomeni riducono la concentrazione di ioni idrogeno, ossia l'acidità, dell'acqua che giunge al lago o al fiume. Gli acidi possono inoltre dissolvere i minerali delle argille presenti nel suolo liberando alluminio, dannoso per le piante e la vita acquatica.

Le foreste morte o morenti sugli alti versanti delle Green Mountains nel Vermont appaiono in rosso in questa immagine a falsi colori, ripresa da satellite con una pellicola sensibile all'infrarosso, che permette di distinguere la clorofilla (*a sinistra*). Nella fotografia scattata a Whiteface Mountain, nello Stato di New York (*a destra*), si vedono parecchi abeti morti. Molti studiosi ritengono che le piogge acide, forse congiuntamente ad altri inquinanti, abbiano provocato il rapido declino di alcune foreste di tipo alpino negli Stati Uniti orientali, sebbene non si sia potuto accertare un rapporto di causa ed effetto. L'immagine ripresa dal *Thematic Mapper* del Landsat è stata fornita da James E. Vogelmann dell'Università del New Hampshire; la fotografia è di Ann Carey del Forest Service statunitense.

sono diminuiti negli ultimi 15 anni grazie alle misure di controllo dell'inquinamento previste dal Clean Air Act, emanato nel 1975, e da ulteriori leggi in proposito approvate dai singoli Stati. Anche il tasso di acidificazione dei laghi nella regione sembra essere diminuito. Per invertire effettivamente la tendenza, tuttavia, le deposizioni acide dovranno essere ridotte molto più drasticamente e parecchi politici e studiosi si stanno chiedendo fino a che punto sarà necessario spingersi e quanto velocemente occorrerà agire.

Per rispondere con esattezza a queste domande dobbiamo sapere fino a quando i processi che avvengono nel suolo potranno esercitare un effetto tampone o «intrappolare» l'acidità nelle regioni minacciate e con quale rapidità i laghi possono ripristinare la loro capacità di neutralizzazione. Dobbiamo anche comprendere i rapporti tra piogge acide e declino delle foreste. Alcune delle risposte dovrebbero essere fornite dal rapporto del NAPAP previsto per il 1990; alcuni studiosi hanno già ipotizzato, tuttavia, che per proteggere laghi e corsi d'acqua in zone a rischio come i monti Adirondack sarà necessario ridurre le deposizioni acide a meno del 50 per cento del livello attuale.

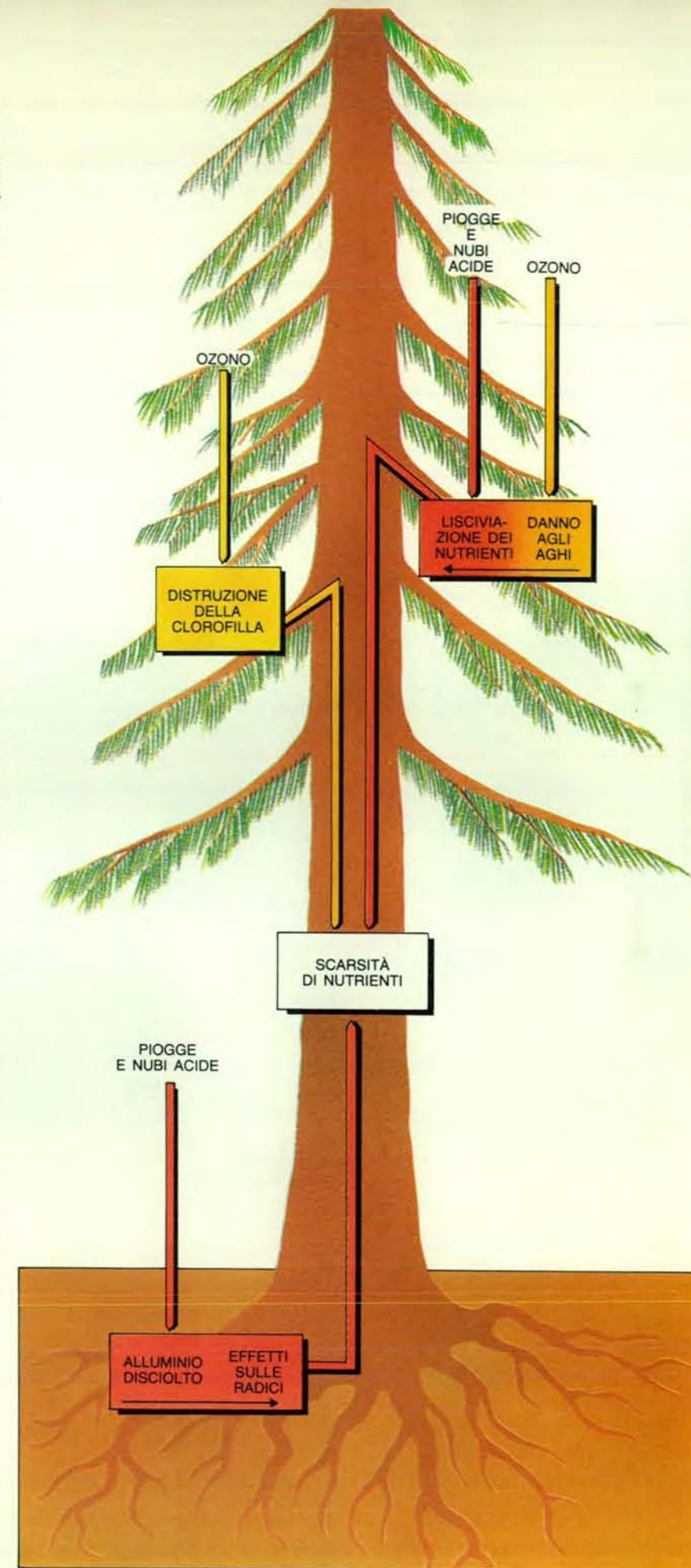
Dove e di quanto dovranno essere ridotte le emissioni per raggiungere questo obiettivo? Alcune linee guida saranno fornite da due modelli al calcolatore, attualmente in fase sperimentale, destinati a valutare la produzione, il trasporto e la deposizione di sostanze acide su vasta scala: il Regional Acid Deposition Model (RADM), promosso dalla US Environmental Protection Agency, e l'Acid Deposition and Oxidant Model (ADOM), sostenuto da Enti governativi canadesi e tedesco-occidentali. I modelli prendono in considerazione tutti i processi chimici e meteorologici che avvengono nell'atmosfera e sicuramente agiscono su molecole contenenti zolfo, azoto e carbonio. (Si tiene conto anche delle molecole contenenti carbonio perché hanno un ruolo nel produrre gli ossidanti che trasformano le emissioni di zolfo e azoto in acidi.)

Data una serie di posizioni di sorgenti, di livelli di emissioni e di condizioni atmosferiche, questi modelli possono prevedere i fenomeni meteorologici e il chimismo dell'atmosfera per consentire di dedurre, con una risoluzione geografica di circa 130 chilometri quadrati o meno, la quantità di deposizioni acide su un'intera regione nel corso di 1-4 giorni. Facendo la media dei risultati calcolati per svariate condizioni atmosferiche, i modelli permettono anche di prevedere l'andamento a lungo termine delle deposizioni per un determinato andamento delle emissioni, il che dovrebbe renderli uno strumento prezioso nel contesto di una strategia per la riduzione delle emissioni.

In che modo si possono controllare gli inquinanti che danno origine alle piogge acide? Il sistema più diretto sarebbe quello di utilizzare minori quantità di combustibili fossili per i trasporti e per la produzione di energia. L'incremento dei mezzi di trasporto di massa e la costruzione di autovetture con consumi di carburante più ridotti possono diminuire l'uso di combustibili nel settore dei trasporti, ma la produzione di energia rappresenta un problema più difficile. Nonostante lo sviluppo di migliori strategie per il risparmio energetico, è probabile che a lungo termine i consumi siano destinati ad aumentare e le attuali alternative alle centrali a combustibile fossile non appaiono molto promettenti. La produzione di energia idroelettrica è limitata dalla scarsità di siti adatti, mentre quella di energia nucleare si trova a fare i conti con problemi economici e con la scarsa fiducia del pubblico nella sua sicurezza.

Il fattore chiave per il controllo delle piogge acide, quindi, è necessariamente la riduzione delle emissioni dalle centrali a combustibili fossili, in particolare quelle a carbone. La metodologia che ha già condotto a una riduzione delle emissioni di composti dello zolfo negli Stati Uniti, in Germania Federale e in Giappone combina l'uso di carbone con un tenore di zolfo naturalmente basso, o che è stato sottoposto a lavaggio per rimuoverne zolfo e altri contaminanti, con la desolforazione dei gas di scarico. Questa consiste nello spruzzamento dei gas di scarico caldi con una fanghiglia calcarea umida che rimuove fino al 90 per cento del biossido di zolfo. Tuttavia può essere difficile disfarsi dei fanghi residui contenenti zolfo e il desolforatore riduce l'efficienza di una centrale, causando un aumento del consumo di carbone di parecchi punti percentuali a parità di potenza generata. Inoltre il processo non riduce per nulla l'emissione di ossidi di azoto.

Le tecnologie innovative per le centrali elettriche sviluppate congiuntamente dal Governo federale e dall'industria nell'ambito del Clean Coal Demonstration Program, approvato nel 1984, offrono una soluzione di maggiore portata. Tre tecnologie per la combustione «pulita» del carbone vengono già utilizzate in grandi centrali dimostrative (si veda l'articolo *Centrali a carbone per il futuro* di Richard E. Balzhiser e Kurt E. Yaeger in «Le Scienze» n. 231, novembre 1987). Nel sistema denominato «combustione a letto fluido», il carbone frantumato e mescolato con calcare minuto viene tenuto in sospensione da un'intensa corrente d'aria ascendente; la zona in cui avviene la combustione è percorsa da tubi di caldaia che convogliano vapore alle turbine della centrale. La miscelazione turbolenta di carbone e aria permette alla combustione di avvenire a una

Le piogge acide e l'ozono potrebbero contribuire a creare una condizione di scarsità di apporti nutritivi in una conifera, secondo un modello, attualmente considerato assai plausibile, del loro ruolo nel declino delle foreste. L'ozono potrebbe agire sia distruggendo la clorofilla (indispensabile per la fotosintesi) sia degradando il rivestimento ceroso degli aghi. Le piogge acide o le goccioline ancora più acide delle nubi potrebbero allora penetrare più facilmente nel tessuto degli aghi lisciviando le sostanze nutritive in esso contenute. Nel suolo gli acidi potrebbero aggravare la carenza di sostanze nutritive mobilizzando l'alluminio, che è in grado di spostare il calcio dai suoi siti di legame sui peli radicali. Sottoposto al trauma di un apporto troppo scarso di sostanze nutritive, l'albero sarebbe facilmente attaccato da insetti, malattie e altri agenti.

temperatura più bassa e più uniforme che in una caldaia di tipo convenzionale, il che riduce la formazione di ossidi di azoto. Inoltre il calcare assorbe in modo efficiente il biossido di zolfo. In una variante di questa tecnologia, la combustione a letto fluido sotto pressione, il carbone è bruciato in aria compressa, migliorando in tal modo l'efficienza della centrale.

Nella terza tecnologia, che integra la gassificazione con il ciclo combinato, il carbone viene fatto reagire con vapore e aria ad alta temperatura, producendo un gas composto principalmente di idrogeno e monossido di carbonio. La combustione del gas di carbone permette di azionare direttamente una turbina; il calore residuo all'uscita delle turbine a gas può essere recuperato per generare vapore, che a sua volta aziona altre turbine per la produzione di elettricità. Una centrale di questo tipo ha un'efficienza molto maggiore rispetto a una convenzionale ed emette quantità nettamente più ridotte di biossido di zolfo e ossidi di azoto.

Dotare di desolforatori gli impianti esistenti è il modo più rapido per ridurre le emissioni delle centrali termoelettriche. Quasi la metà degli impianti funzionanti a carbone degli Stati Uniti è stata costruita prima del 1975 e non è dotata di sistemi per la limitazione delle emissioni inquinanti. Concentrate nella parte orientale degli Stati Uniti, queste centrali sono responsabili di gran parte delle emissioni di zolfo in tutto il paese. L'aggiunta di desolforatori convenzionali potrebbe ridurre le emissioni di biossido di zolfo da tutte le centrali a meno di metà del livello attuale e la riduzione potrebbe aver luogo entro 15 anni; tuttavia le emissioni di ossidi di azoto rimarrebbero invariate. Inoltre le aziende elettriche si oppongono alle spese necessarie per l'installazione e il funzionamento dei desolforatori e alla perdita di efficienza che ne deriverebbe.

Le riduzioni prevedibili nelle emissioni annuali di biossido di zolfo (*in alto*) e di ossidi di azoto (*in basso*) dalle centrali elettriche funzionanti a combustibile fossile variano in dipendenza delle scelte tecnologiche. Sostituire gli impianti che raggiungono i 50 anni di vita con nuove centrali dotate di desolforatori dei gas di scarico diminuirebbe solo il biossido di zolfo e la riduzione avrebbe luogo lentamente (*in rosso*). Se tutte le centrali esistenti venissero munite di desolforatori entro 15 anni, la diminuzione sarebbe molto più netta, ma riguarderebbe ancora solo il biossido di zolfo (*in blu*). In entrambi i casi le emissioni di ossidi di azoto continuerebbero ad aumentare con la costruzione di nuove centrali per soddisfare il crescente fabbisogno di energia elettrica. La ristrutturazione delle centrali costruite tra il 1955 e il 1975, con dotazione di apparecchiature per tecnologie «pulite» come la combustione del carbone in letto fluido, determinerebbe la diminuzione più drastica nei livelli di entrambi gli inquinanti. Le fasce in verde mostrano gli intervalli previsti per le riduzioni nel caso le centrali venissero ristrutturate dopo 40-50 anni dalla loro costruzione.

Le tecnologie per la combustione pulita del carbone rappresentano un'interessante alternativa. Gli sforzi per il controllo delle precipitazioni acide devono essere indirizzati sulle centrali più vecchie, molte delle quali potrebbero entro breve tempo essere messe fuori servizio oppure ristrutturate. La loro sostituzione con nuove centrali convenzionali fornite di desolforatori consentirebbe soltanto una riduzione modesta delle emissioni e i costi per la progettazione, la costruzione e l'ottenimento delle autorizzazioni richieste sarebbero insostenibili. Invece gran parte delle vecchie centrali (i 410 impianti costruiti fra il 1955 e il 1975) potrebbero essere fornite di una nuova sezione di combustione che incorpori una delle tecnologie per la combustione pulita.

Una centrale così ristrutturata potrebbe conservare in gran parte le sue attuali apparecchiature per il trattamento del carbone e delle ceneri e gli strumenti per il ciclo del vapore e la generazione di elettricità. Di conseguenza, l'adeguamento di una centrale esistente sarebbe veloce ed economico rispetto alla costruzione di un nuovo impianto e presenterebbe un ulteriore vantaggio per le aziende elettriche: le nuove apparecchiature potrebbero essere aggiunte per gradi, il che permetterebbe di adeguare la capacità di produzione all'effettiva richiesta di energia.

La ristrutturazione di vecchie centrali prospetta in definitiva le maggiori riduzioni nelle emissioni, che riguarderanno l'intera gamma di inquinanti coinvolti nel fenomeno delle precipitazioni acide. Questa strategia potrebbe diminuire le emissioni di biossido di zolfo di più dell'80 per cento e quelle di ossidi di azoto di più del 50 per cento; una ridotta emissione di composti dell'azoto diminuirebbe a sua volta la formazione di ozono nella troposfera.

Sappiamo ora che il termine «precipitazioni acide» comprende una gran varietà di fenomeni. Gli ossidi di azoto, per esempio, influenzano il ciclo chimico che trasforma il biossido di zolfo in acido solforico e l'ozono che essi contribuiscono a produrre opera forse di concerto con le piogge acide nella distruzione delle foreste. Lo ione nitrato e l'acidità che gli è associata possono danneggiare non solo gli ecosistemi terrestri ma, secondo recenti studi, anche quelli delle acque costiere. La riduzione nelle emissioni che potrà aversi con l'adeguamento delle centrali potrebbe rendere meno drammatici tutti questi effetti.

L'inconveniente è che queste riduzioni, per quanto notevolissime, avverranno necessariamente con lentezza. Il recente declino del contenuto di zolfo dell'atmosfera e delle precipitazioni nella regione nordorientale degli Stati Uniti e il rallentamento dell'acidificazione dei laghi fanno pensare che vi sia ancora spazio per l'intervento. Possiamo probabilmente rinunciare alla soluzione a breve termine, che consiste nel dotare le centrali esistenti di impianti per la limitazione dell'inquinamento, in favore dell'impostazione più ampia e più economica che prevede la ristrutturazione degli impianti. Tuttavia, per accelerare la trasformazione, è necessario che il Governo intervenga ponendo scadenze precise. Se le aziende si limitassero ad adeguare le centrali solo quando ne sorgesse la necessità (per esempio dopo circa 50 anni dalla loro costruzione) il processo sarebbe completato solo molto dopo l'inizio del prossimo secolo.

La tecnologia ha superato la scienza e ci ha offerto un'opzione per affrontare il problema delle precipitazioni acide che potrebbe risultare vantaggiosa, quale che sia la soluzione dei restanti dubbi scientifici. Il bisogno urgente di ridurre l'interferenza umana nel complesso chimismo della biosfera è già drammaticamente chiaro.

Gli studi sui fenomeni di acidificazione in Italia

Anche in Italia, come nella maggior parte dei paesi industrializzati, si è avuto negli ultimi decenni un sensibile aumento dell'acidità atmosferica. Se ne avvertono gli effetti sulle caratteristiche chimiche delle deposizioni atmosferiche e delle acque superficiali, sullo stato di salute della vegetazione e sul deperimento del patrimonio monumentale.

Per quanto riguarda il primo aspetto, studi eseguiti da vari istituti hanno evidenziato che le deposizioni acide interessano in maniera cronica gran parte del Piemonte, l'area tributaria del lago Maggiore con il Canton Ticino (Svizzera) e la Lombardia occidentale. I valori medi annui di *p*H sono generalmente compresi tra 4,2 e 4,4, con minimi di 4,11 nella città di Torino. Si tratta dunque di valori molto simili a quello riportato da Volker A. Mohnen come media del nord-est degli Stati Uniti (4,2) che è, ricordiamolo, un'area fra le più densamente abitate e industrializzate del mondo. Tornando alla situazione italiana, in altre località di prelievo in Lombardia, Veneto ed Emilia-Romagna le deposizioni presentano un'acidità minore o trascurabile, ma le alte concentrazioni di solfati, nitrati e ammonio testimoniano l'elevato livello di emissioni di inquinanti in atmosfera. Infine altri studi hanno documentato episodi di deposizione acida in Toscana, Marche, Campania, Sicilia e Sardegna.

Questi risultati sono però confrontabili qualitativamente e non quantitativamente, a causa delle diverse metodologie usate per il prelievo e l'analisi dei campioni. Il Ministero dell'ambiente, con il supporto scientifico dell'Istituto italiano di idrobiologia del CNR, da circa un anno svolge un'azione di coordinamento fra le varie iniziative di studio per ricavare un quadro complessivo della situazione della chimica delle deposizioni in Italia. Il coordinamento si esplica nell'individuazione di una metodologia comune per il prelievo e il trattamento dei campioni, nell'organizzazione di intercalibrazioni per assicurare la confrontabilità dei dati chimici, nel raccogliere in una banca dati i risultati per la successiva elaborazione con metodiche unificate.

Di concerto con questa attività, il CNR, a continuazione delle ricerche in corso, ha varato un progetto strategico per approfondire le metodologie di studio di questi problemi. Il progetto si articola in tre aree, riguardanti la deposizione umida, la deposizione secca e gli aspetti strumentali. Particolare cura sarà posta per assicurare metodologie uniformi e risultati confrontabili con quelli prodotti in campo internazionale.

Passando all'effetto sulle acque superficiali, la situazione italiana non è fortunatamente drammatica come quella descritta da Mohnen per i laghi della Scandinavia e del nord-est degli Stati Uniti. Uno studio svolto dall'Istituto italiano di idrobiologia, con il concorso finanziario del CNR, della CEE e dell'ENEL, ha considerato circa 650 laghi italiani, dei quali 540 situati sulle Alpi, allo scopo di quantificare i livelli di acidificabilità (definita come possibilità di un'acqua lacustre di neutralizzare apporti acidi) e di acidificazione (definita come perdita di alcalinità nel tempo) dei laghi italiani. Riguardo all'acidificabilità, grazie alla diffusa presenza di rocce carbonatiche gran parte dei laghi di bassa quota, fra cui quelli di maggiori dimensioni (Maggiore, Como, Lugano, Iseo e Garda) non risultano a rischio di acidificazione. Al contrario, oltre il 50 per cento dei laghi alpini risulta a rischio, come evidenziato dalle basse concentrazioni di bicarbonati. Il livello di acidificazione è stato valutato con un modello applicabile solo a una parte dei laghi alpini; si è misurata una perdita di circa 50 microequivalenti/litro di bicarbonati rispetto alla concentrazione originaria.

Gli effetti delle deposizioni acide sulla vegetazione sono stati valutati su scala nazionale dal Corpo forestale dello Stato e su scala regionale da istituti universitari e dalla Società botanica italiana. Il rilevamento nazionale ha evidenziato un livello di alterazione medio del 10 per cento. Tale dato è molto più basso di quelli rilevati dai competenti organi svizzeri e austriaci per le aree di confine. Si può citare il 65 per cento di deterioramento rilevato dall'Istituto federale di ricerche forestali per il Canton Ticino, area politicamente svizzera ma geograficamente italiana. Queste incongruità sottolineano l'urgenza dell'adozione di metodologie uniformi a cui si è accennato in precedenza. Per quanto riguarda il patrimonio monumentale, tutti gli studi eseguiti sono concordi nell'indicare nell'inquinamento atmosferico le cause dell'accelerato deterioramento degli ultimi decenni.

Difficile stimare le perdite economiche dovute agli effetti delle rideposizioni al suolo degli inquinanti atmosferici. Tuttavia, nei paesi dove tale stima è stata fatta, i totali superavano di gran lunga le cifre richieste per ridurre in misura significativa l'inquinamento atmosferico. E, al di là degli aspetti economici, considerazioni etiche e di buon senso e urgenze di tipo ambientale concordano nell'indicare come esigenza non più differibile una drastica riduzione delle emissioni di inquinanti in atmosfera.

Rosario Mosello
Istituto italiano di idrobiologia del CNR
Pallanza

BIBLIOGRAFIA

LIKENS GENE E., WRIGHT RICHARD F., GALLOWAY JAMES F. e BUTLER THOMAS J., *Pioggia acida* in «Le Scienze» n. 136, dicembre 1979.

Special Issue on the Experimental Lakes Area in «Canadian Journal of Fisheries and Aquatic Science/Journal canadien des sciences halieutiques et aquatiques», 21, n. 37, marzo 1980.

COWLING ELLIS B., *Acid Precipitation in Historical Perspective* in «Environmental Science and Technology», 16, n. 2, febbraio 1982.

WOODMAN JAMES N. e COWLING ELLIS B., *Airborne Chemicals and Forest Health* in «Environmental Science and Technology», 21, n. 2, febbraio 1987.

MOSELLO ROSARIO e DE MARGARITIS BRUNA (a cura), *Deposizioni acide: un problema per acque e foreste*, Consiglio nazionale delle ricerche - Istituto italiano di idrobiologia, Verbania Pallanza, 1987.

Oltre verità e bellezza: una quarta famiglia di particelle

Recenti esperimenti con gli acceleratori di particelle e considerazioni cosmologiche indicano l'esistenza di un'altra famiglia di quark e di leptoni oltre alle tre note, ma fanno pensare che non ve ne siano più di cinque

di David B. Cline

Gli scienziati che studiano la natura ultima della materia sono convinti che la varietà dell'universo fisico si possa spiegare supponendo l'esistenza di poche particelle fondamentali. Tale convinzione è stata messa a dura prova. Attorno alla metà di questo secolo, la semplicità emergente del protone, del neutrone, dell'elettrone e delle loro corrispondenti antiparticelle si dissolse in centinaia di particelle subnucleari. Negli anni settanta tale semplicità parve riemergere con la scoperta del quark, ma la matassa si ingarbugliò nuovamente allorché vennero individuati parecchi altri quark.

Oggi sembra che la contesa si possa appianare con un accordo di compromesso. Da un lato le osservazioni di alcuni isotopi, quali il deuterio e gli isotopi dell'elio presenti nello spazio profondo, congiuntamente a esperimenti di laboratorio compiuti con acceleratori indicano che il numero di particelle fondamentali è veramente limitato. Dall'altro lato, alcuni indizi fanno pensare che questo numero possa comprendere più delle tre famiglie di quark di cui è oggi nota l'esistenza. Ad aumentare l'entusiasmo vi è la prospettiva che risposte a quesiti fondamentali, quale quello sull'origine della massa, possano essere alla portata dei nuovi sofisticati acceleratori che iniziano oggi a entrare in funzione.

Per capire perché alcuni fisici ritengono che possa esistere una quarta famiglia di quark, ma che non ve ne siano molte più di quattro, si deve prima considerare ciò che attualmente viene o non viene spiegato dal modello standard della fisica delle particelle. Quasi ogni disciplina ha il suo modello standard; quello della fisica delle particelle si basa sull'ipotesi che la materia comune sia composta da due tipi di particelle, i quark e i leptoni, e che le forze tra di essi vengano mediate da una terza categoria di particelle, i bosoni. I leptoni comprendono l'elettrone e il neutrino, mentre i quark si combinano in particelle più grandi quali il protone e il neutrone. Un ben noto esempio di bosone è il fotone, che media la forza elettromagnetica.

Sono state scoperte sperimentalmente tre famiglie di quark, ciascuna comprendente due particelle, per un totale di sei quark. La prima famiglia comprende i quark «su» (*up*) e «giù» (*down*). Il quark su ha una massa di circa 4 milioni di elettronvolt (MeV), ovvero circa 1/250 della massa del protone (che è prossima a un miliardo di elettronvolt, ossia 1 GeV). La massa del quark giù è leggermente superiore, pari a circa 7 MeV. La seconda famiglia è costituita dal quark «strano» (*strange*) e dal quark «incantato» (*charm*), con masse di circa 150 e 1300 MeV rispettivamente. La terza famiglia comprende il quark «basso» (*bottom*), noto anche come «bellezza» (*beauty*), con una massa di 5,5 GeV, e il quark «alto» (*top*), o «verità» (*truth*), che non è stato ancora individuato (*si veda l'illustrazione a pagina 30 in basso*).

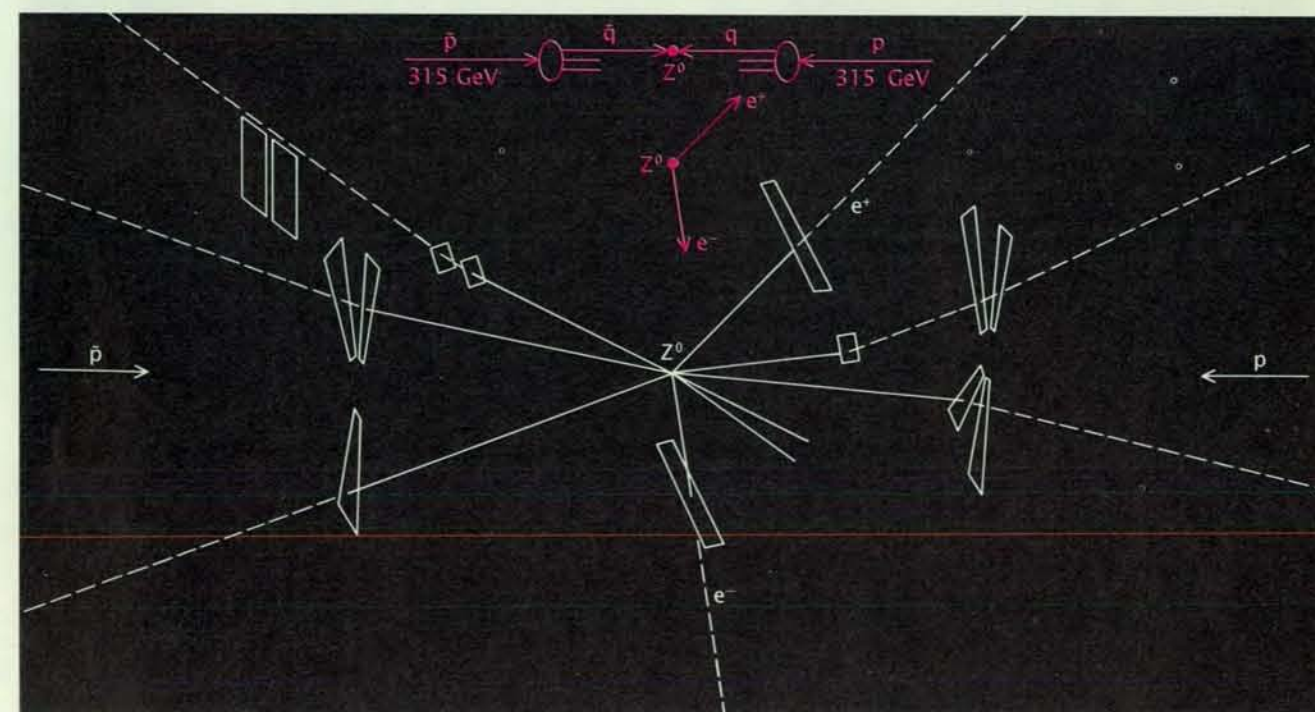
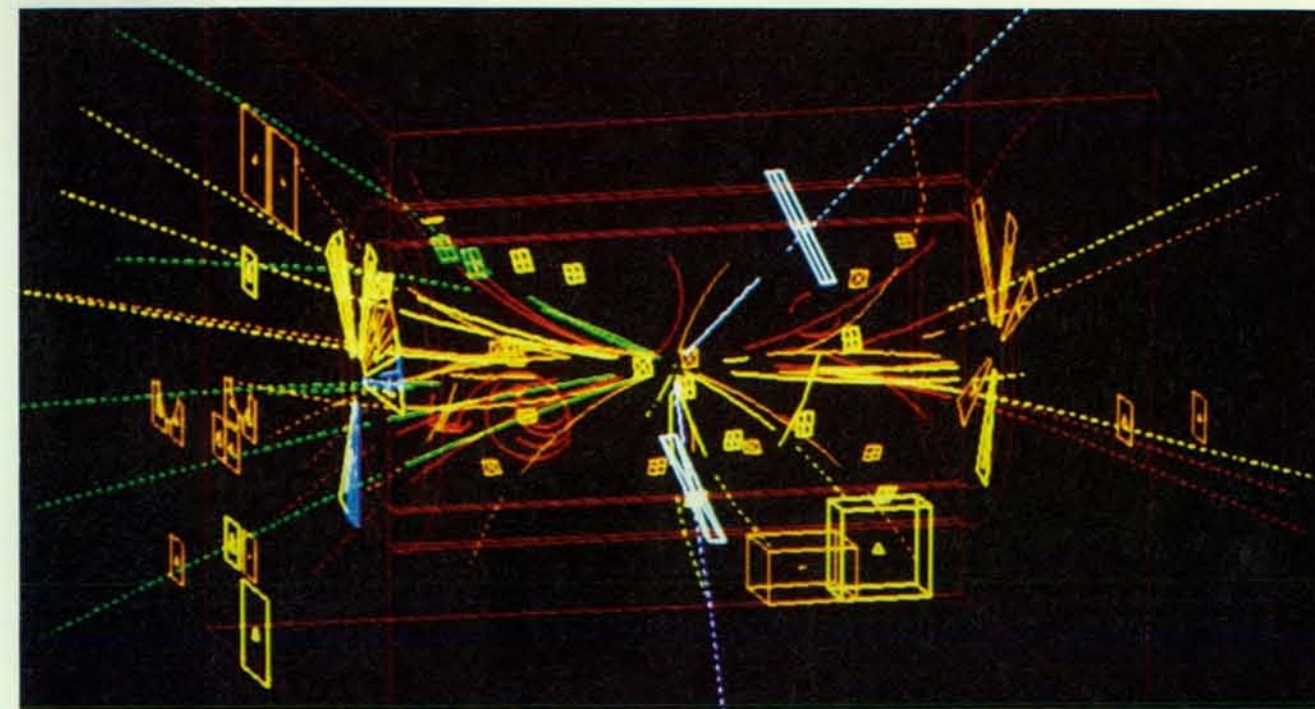
Mentre il protone ha una carica elettrica positiva unitaria, i quark hanno una carica frazionaria. Il quark su ha una carica frazionaria 2/3, mentre il quark giù ha carica −1/3. Dato che il protone è formato da due quark su e da un quark giù, la carica totale, come richiesto, è 1. Il neutrone è composto da due quark giù e da un quark su, che danno una carica elettrica totale nulla. Allo stesso modo si possono combinare i vari quark e formare tutte le altre particelle note che non siano leptoni o bosoni.

I quark di ogni famiglia sono circa 10 volte più pesanti di quelli della famiglia precedente. Questo fatto fa pensare che ogni nuovo quark debba essere molto pesante. In effetti, recenti esperimenti condotti presso il Laboratorio europeo per la fisica delle particelle del CERN (l'Organizzazione europea per la ricerca nucleare) fissano un limite inferiore di circa 50 GeV per la massa del quark verità, non ancora scoperto. In ciascuna famiglia, però, le masse dei quark sono comprese entro un ordine di grandezza e quindi si prevede che il quark verità abbia una massa non più di 10 volte maggiore di quella del quark bellezza. (Se futuri esperimenti con acceleratori produrranno un quark verità molto pesante, i fisici teorici si troveranno di fronte a un grosso rompicapo.)

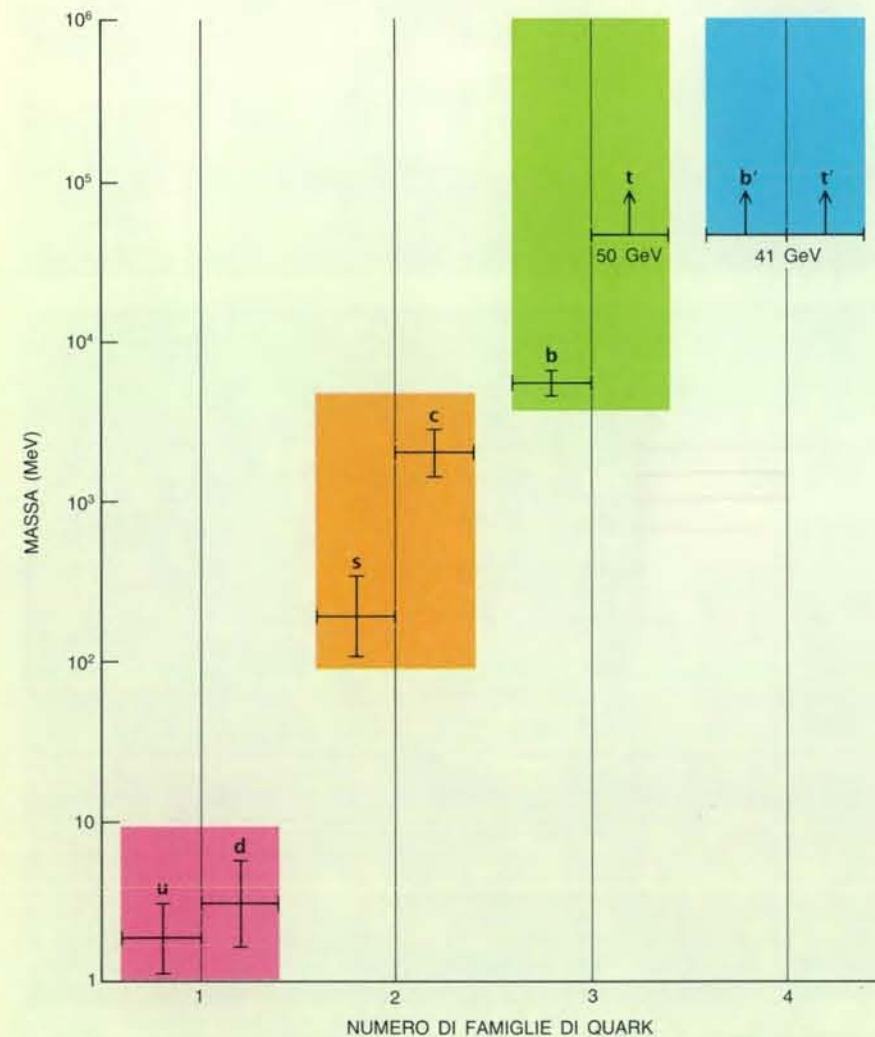
Le famiglie di leptoni

Si è visto sperimentalmente che a ogni famiglia di quark è associata una famiglia di leptoni, formata sempre da un leptone carico e da un leptone neutro. Nella prima famiglia il leptone carico è l'elettrone e quello neutro è il neutrino elettronico; nella seconda famiglia i leptoni sono il muone e il neutrino muonico, mentre nella terza sono la particella tau e il neutrino tauonico.

Rispetto ai quark i leptoni sono estremamente leggeri (*si veda l'illustrazione a pagina 31*). L'elettrone ha una massa di circa 1/2 MeV; le osservazioni di neutrini elettronici emessi dalla supernova 1987A pongono un limite superiore di 16 elettronvolt alla loro massa. Si è giunti a questa conclusione in base alla considerazione che la velocità di particelle con massa di riposo non nulla varia con la loro energia; i tempi di arrivo di un fiotto di neutrini pesanti dovrebbero quindi distribuirsi in un periodo finito. Il fatto che tutti i neutrini provenienti dalla supernova 1987A siano arrivati sulla Terra entro 13 secondi l'uno dall'altro porta al limite di 16 eV. E, dal momento che questo è soltanto un limite superiore, la vera massa del neutrino elettronico potrebbe essere nulla.

Il leptone carico della seconda famiglia, il muone, è circa 200 volte più pesante dell'elettrone, ma sotto tutti gli altri aspetti è a esso identico. I limiti sperimentali impongono che la massa del neutrino muonico sia inferiore a circa 100 000 eV. I limiti cosmologici sono però molto più restrittivi in quanto stabiliscono che la massa di ogni neutrino debba essere inferiore al valore estremamente basso di 65 eV. Il neutrino tauonico non è stato ancora osservato direttamente, ma la sua compagna, la particella tau, è stata scoperta nel 1976 da Martin L. Perl e collaboratori presso lo Stanford Linear Accelerator Center (SLAC). La particella ha una massa di 1,8 GeV. Dato che, secondo la teoria, la particella tau deve avere un proprio neu-

La produzione e il decadimento di un bosone vettore intermedio neutro (la particella Z^0) sono mostrati in un'immagine al calcolatore e in uno schema esplicativo. La particella Z^0 è stata prodotta da collisioni quark-antiquark nell'esperimento condotto all'UA1 (Underground Area 1) con il collisore protone-antiprotone del Laboratorio europeo per la fisica delle particelle del CERN; la particella decade in una coppia elettrone-positrone. (Si notano tracce di altre particelle formate nella collisione dei fasci.) La particella Z^0 può decadere in tutte le famiglie di leptoni esistenti, neutrini compresi. Se vi fossero infinite famiglie di neutrini non si osserverebbe mai il decadimento della Z^0 nella coppia elettrone-positrone. Il fatto che venga osservato limita probabilmente il numero di famiglie a cinque.

FAMIGLIA	1	2	3	NON ANCORA RIVELATA 4
LEPTONI	ν_e e^-	ν_μ μ^-	ν_τ τ^-	ν_L L
QUARK	u d	c s	t b	t' b'
BOSONI DI GAUGE: FOTONE, GLUONE, $W^{\pm}$, Z^0, GRAVITONE				SCALARE DI HIGGS

In base al modello standard della fisica delle particelle, la materia sarebbe formata da quark e leptoni e le forze fra le particelle sarebbero mediate da bosoni. Ogni famiglia di leptoni è costituita da un leptone carico e da un leptone neutro assai più leggero, nell'ordine: l'elettrone (e^-) e il neutrino elettronico (ν_e), il muone (μ) e il neutrino muonico (ν_μ), la particella tau (τ) e il neutrino tauonico (ν_τ). Secondo il modello standard ogni famiglia di leptoni è associata a una famiglia di quark formata a sua volta da due particelle: un quark «su» (*up* o *u*) e uno «giù» (*down* o *d*); un quark «strano» (*strange* o *s*) e uno «incantato» (*charm* o *c*); un quark «basso» (*bottom* o *b*), detto anche «bellezza» (*beauty*) e uno «alto» (*top* o *t*), detto anche «verità» (*truth*). Quest'ultimo non è stato ancora rivelato. Se la corrispondenza tra le famiglie continua in una quarta generazione, si dovrebbero scoprire un quarto leptone carico (*L*), un quarto neutrino (ν_L) e altri due quark, indicati con *t'* e *b'*.

Le masse dei cinque quark attualmente conosciuti sono espresse in milioni di elettronvolt (MeV). Le masse dei quark che appartengono a famiglie successive differiscono di circa un ordine di grandezza. Recenti esperimenti impongono un limite inferiore per la massa del quark verità pari a 50 GeV (miliardi di elettronvolt); si possono ipotizzare limiti anche per le masse dei quark della quarta famiglia, che dovrebbero essere superiori a 41 GeV.

trino, i fisici sono convinti dell'esistenza del neutrino tauonico. Esperimenti diretti indicano che la sua massa è inferiore a 70 MeV; ancora una volta il limite cosmologico sembra invece essere al di sotto di 65 eV.

Come le famiglie di quark, anche le famiglie di leptoni sono raggruppate in intervalli di massa diversi. La massa del muone è di circa due ordini di grandezza maggiore di quella dell'elettrone, mentre la particella tau è a sua volta circa 20 volte più pesante del muone. Ci si potrebbe quindi aspettare che qualsiasi ulteriore leptone carico abbia una massa prossima a 40 GeV. Come dirò in seguito, gli attuali limiti inferiori sperimentali alla massa di nuovi leptoni carichi sono in accordo con questa previsione. Anche le masse dei neutrini possono essere molto distanziate, ma, essendo stati fissati solo limiti superiori, tutto ciò che si può dire oggi è che la massa dei neutrini è molto piccola in confronto alla massa dei rispettivi leptoni carichi.

L'esiguità della massa dei neutrini porta alla seconda importante differenza tra quark e leptoni: oltre al fatto che i leptoni sono molto più leggeri in termini assoluti rispetto ai quark loro associati, i rapporti di massa all'interno delle famiglie di quark sono di gran lunga inferiori a quelli all'interno delle famiglie di leptoni associate. In ciascuna famiglia di quark il rapporto tra le masse dei quark non è mai superiore a circa 10. Per esempio, il rapporto tra la massa del quark giù e quella del quark su è circa due. I leptoni presentano un quadro completamente diverso: considerando validi i limiti superiori alla massa dei neutrini, il rapporto tra la massa dell'elettrone e quella del neutrino elettronico è di circa 10 000; se si dovesse accertare che il neutrino ha massa nulla, tale rapporto sarebbe infinito.

Il modello standard ha avuto successo anche nella descrizione dei bosoni, le particelle che mediano le forze tra altre particelle. Nella teoria dell'elettromagnetismo di Maxwell questo ruolo è interpretato dal fotone, il portatore della forza elettromagnetica. L'attuale modello standard comprende anche la forza debole, che determina il decadimento radioattivo, e la forza forte, che tiene legato il nucleo. Sono quindi necessari altri bosoni. Le interazioni deboli (come il decadimento di un neutrone in un protone e in un elettrone) che prevedono lo scambio di carica elettrica sono mediate dal cosiddetto bosone vettore intermedio carico o *W*. Altre interazioni deboli che non richiedono uno scambio di carica sono mediate dal bosone vettore intermedio neutro o Z^0. Uno dei grandi trionfi del modello standard è stata la previsione della massa delle particelle *W* e Z^0. Entrambe le particelle sono state poi scoperte al CERN nel 1983 e avevano la massa prevista. Anche tutte le altre proprietà osservate delle particelle sono in notevole accordo con la teoria.

Difetti del modello

Nonostante tali successi, il modello standard presenta un certo numero di seri difetti. Per prima cosa, esso non fa alcuna previsione sul numero di famiglie di quark e di leptoni. Perché esistono almeno tre famiglie, se per formare i comuni protoni, neutroni ed elettroni presenti nell'universo basta la prima famiglia? O, come disse I. I. Rabi 50 anni fa: «Il muone, chi l'ha chiesto?»

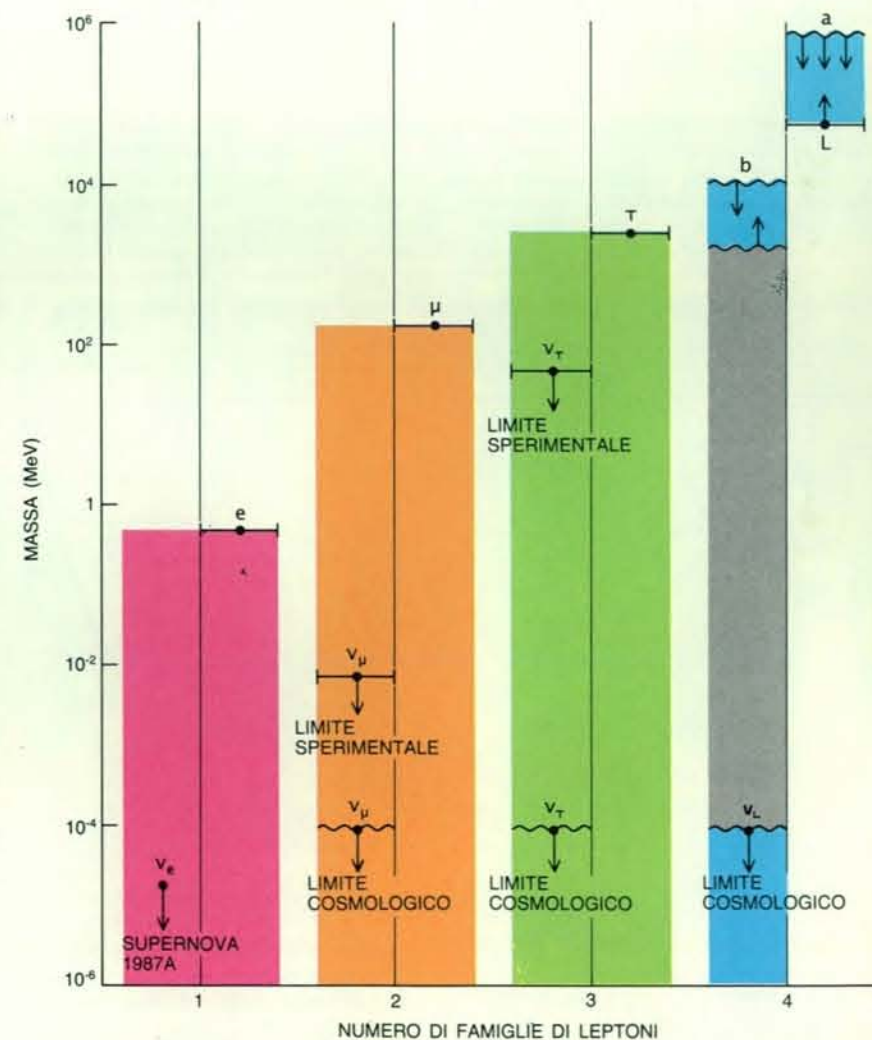
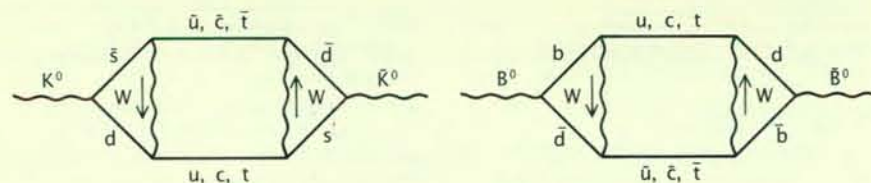
Il modello standard inoltre non è in grado di prevedere la massa di tutte le restanti particelle; il limite inferiore di 50 GeV per la massa del quark verità è un risultato sperimentale e nessuno sa quale sia il limite superiore. Il modello non spiega nemmeno la gerarchia delle masse dei quark e dei leptoni descritta in precedenza. Perché le famiglie di quark sono separate, per quanto riguarda la massa, all'incirca da un ordine di grandezza e quelle di leptoni da due ordini di grandezza? Perché i rapporti tra le masse dei quark all'interno di una famiglia sono così piccoli e quelli tra le masse dei leptoni così grandi? Sono stati compiuti negli anni molti tentativi di interpretazione numerologica per spiegare questa distribuzione di massa, ma nessuno ha avuto successo, cosicché essa rimane uno dei grandi inspiegati perché del modello standard.

Le domande relative al numero di famiglie, alla massa e alla gerarchia delle masse non sono le sole alle quali il modello standard non è in grado di rispondere. Un altro grande mistero insoluto è il fatto che, mentre si osservano spesso i diversi tipi di quark trasformarsi uno nell'altro, non si è mai osservata la stessa cosa per i leptoni. Per esempio, il quark incantato può decadere in un quark strano e in una particella chiamata «*W* virtuale» (che può essere interpretata come una vera particella *W* con una vita tanto breve da non poter essere osservata direttamente). D'altro canto, non si è mai visto decadere un muone in un elettrone e in un fotone e la probabilità che ciò possa avvenire è stata ridotta per via sperimentale a meno di una parte su 100 miliardi.

Questa proliferazione di misteri ha indotto alcuni teorici a sospettare l'esistenza di una quarta famiglia di quark e leptoni. La scoperta di tale famiglia chiarirebbe alcune questioni da lungo tempo in sospeso. Una di esse riguarda il fenomeno chiamato violazione della parità di carica, che è a sua volta correlato con le probabilità della transizione quark-quark appena esaminata. Fino agli anni sessanta i fisici erano convinti che le proprietà misurabili di qualsiasi sistema fisico dovessero rimanere invariate quando ogni particella è trasformata nella sua antiparticella e il sistema è riflesso in uno specchio. Poiché la trasformazione di una particella nella sua antiparticella richiede un cambiamento di carica e poiché la riflessione speculare è definita inversione di parità, l'enunciato che qualunque sistema debba restare immutato dopo le due operazioni combinate è detto legge dell'invarianza della parità di carica (o invarianza *CP*).

La fiducia riposta nell'invarianza *CP* venne meno nel 1964, allorché Val L. Fitch, James W. Cronin, James H. Christenson e René Turlay della Princeton University indagarono il tasso di decadimento di una particella detta mesone *K* neutro, o kaone. Il kaone decade solitamente in altre tre particelle (per esempio, in tre pioni); una siffatta transizione è in accordo con la legge dell'invarianza *CP*. L'esperimento di Princeton dimostrò però che circa una volta su 500 il kaone decade in due soli pioni, subendo così una transizione che viola l'invarianza *CP*. La conseguenza dell'esperimento fu quindi di respingere quella che veniva ritenuta una legge della natura, anche se la vera origine della violazione della parità di carica nel sistema del kaone rimane ancor oggi inspiegata ed è considerata uno dei grandi misteri della fisica.

Mescolamento di sapori

Per quanto inspiegabile resti la violazione *CP*, la sua entità può essere correlata alle probabilità della transizione quark-quark. I quark si trasformano generalmente in altri membri della propria famiglia, come nel decadimento del quark incantato in un quark strano e in una particella *W* virtuale. Perché avvenga la violazione della parità di carica, i

Lo spettro di massa dei leptoni indica che le masse dei leptoni carichi, come quelle dei quark, sono in ordine gerarchico: la massa della particella tau è di circa un ordine di grandezza superiore a quella del muone e questa è a sua volta superiore di due ordini di grandezza alla massa dell'elettrone. Limiti alla massa del quarto leptone carico (*L*) impongono che essa sia maggiore di 41 GeV. La differenza di massa tra i bosoni vettori *W* e Z^0 fornisce un limite superiore (*a*). Le masse dei tre neutrini non sono note, ma lo studio della supernova 1987A impone che la massa del neutrino elettronico sia inferiore a 16 eV. Il requisito che nessun neutrino sia così pesante da rallentare notevolmente l'espansione dell'universo fissa a 65 eV il limite superiore «cosmologico» per tutti i neutrini. L'area in grigio (*a destra*) è esclusa dalla teoria; il limite *b* è dato dalle ricerche sulla materia oscura.

I diagrammi di Feynman mostrano il «mescolamento di sapori» per il sistema del kaone (K^0, *a sinistra*) e per quello di un mesone *B* (*a destra*). Il mescolamento avviene quando un kaone K^0, che contiene un quark antistrano ($\bar{s}$) e un quark giù (*d*), si trasforma in un antikaone ($\bar{K}^0$): il quark $\bar{s}$ deve trasformarsi in un quark $\bar{d}$ e il quark *d* in un quark *s*. Nel sistema del mesone *B*, formato da un quark *b* e da un quark $\bar{d}$, il quark *b* si deve trasformare in un quark *d* e il quark $\bar{d}$ in un quark $\bar{b}$. Si tratta di transizioni incrociate o a mescolamento di sapori. Il mescolamento è necessario perché si abbia la violazione di parità di carica (*CP*).

quark devono essere in grado di trasformarsi in membri di altre famiglie, un processo definito «mescolamento di sapori» (dato che le famiglie di quark sono misteriosamente caratterizzate dall'avere «sapori» differenti). Si può inoltre dimostrare che un modello standard a due famiglie non sarebbe sufficiente per consentire la violazione *CP* nel sistema del kaone neutro; sono necessarie almeno tre famiglie di quark. In effetti, l'esistenza della violazione della parità di carica è stata la prima prova che ha fatto pensare a una terza famiglia di quark. L'entità della violazione *CP* dipende dalla probabilità con la quale un quark di una famiglia può trasformarsi in un quark di un'altra famiglia, ossia dall'entità del mescolamento.

Una quarta famiglia influenzerebbe l'entità del mescolamento di sapori permettendo un maggior numero di transizioni quark-quark. Recentemente i fisici del Deutsches Elektronen-Synchrotron (DESY) e del CERN hanno scoperto che il mescolamento di sapori nel sistema del mesone *B* è 20 volte maggiore di quello previsto. Il mesone *B* è così chiamato

Esperimenti di mescolamento possono fornire le prove dell'esistenza di una quarta famiglia di quark. I mesoni B^0 compaiono in due varietà, $b\bar{d}$ come nell'illustrazione in alto, o $b\bar{s}$. Entrambi i tipi subiscono il mescolamento di sapori, ovvero si trasformano nelle proprie antiparticelle. Recentemente cinque esperimenti (*curve in colore*), eseguiti al CERN, al DESY, allo SLAC e alla Cornell University, hanno dimostrato che il mescolamento era molto maggiore di quanto previsto. L'area in colore delimitata dai risultati sperimentali indica i valori più probabili del livello di mescolamento per i due tipi di particelle B^0. La maggior parte di tale area si trova al di sopra del limite consentito da un modello standard a tre famiglie e ciò indica che al mescolamento potrebbe contribuire anche una quarta famiglia.

perché è formato da un quark bellezza e da un altro quark, per esempio un quark antigiù. Nel processo di mescolamento il mesone *B* viene trasformato in un anti-*B*, il che richiede che il quark bellezza venga trasformato nel quark giù e il quark antigiù nel quark antibellezza (*si veda l'illustrazione in questa pagina in alto*). Si noti che, come nel sistema del kaone, si tratta di transizioni con mescolamento di sapori. La rapidità con la quale si verifica il mescolamento dipende da tutti i quark esistenti e dalle loro masse; maggiore è il numero di quark, maggiore è l'entità del mescolamento prevista. Il fatto che il mescolamento scoperto al DESY fosse molto maggiore di quello previsto può indicare che a esso abbia contribuito una quarta famiglia di quark. Poiché, tuttavia, la massa del quark verità non è ancora nota, può anche darsi che i risultati si rivelino compatibili con tre famiglie di quark.

Il mescolamento del mesone *B* potrebbe anche far luce sull'origine della violazione della parità di carica stessa, che è stata finora osservata soltanto nel sistema del kaone neutro. Il mescolamento di sapori è una condizione necessaria per la violazione *CP*, ma non è sufficiente. Anche se la violazione *CP* non è ancora stata osservata nei mesoni *B*, l'entità stessa del mescolamento ha convinto alcuni ricercatori che l'osservazione della violazione *CP* nel sistema del mesone *B* possa essere prossima. Se si scoprirà che la violazione *CP* è analogamente grande, è improbabile che il modello standard a tre famiglie sia in grado di spiegarla (a meno che la massa del quark verità non sia inaspettatamente grande), e si dovrà quindi chiamare in causa una quarta famiglia di quark.

Una verifica sperimentale di tale affermazione dovrebbe essere possibile nel prossimo futuro. Le collisioni protone-protone in un acceleratore possono produrre coppie mesone *B*-antimesone *B*, che dovrebbero a loro volta decadere in prodotti comprendenti due leptoni carichi. Questi ultimi potrebbero essere elettroni o positroni. Se la *CP* si conserva, i tassi di decadimento dei mesoni in elettroni e positroni dovrebbero essere uguali, mentre se la *CP* viene violata dovrebbero essere diversi. La verifica sarebbe estremamente sensibile all'esistenza di una quarta famiglia di quark. Un'osservazione della violazione *CP* la metterebbe in relazione univoca con le probabilità di transizione quark-quark e potrebbe condurre all'individuazione dell'origine fondamentale della violazione *CP* in natura.

L'importanza dell'individuazione della violazione della parità di carica nel sistema di decadimento del mesone *B* sta ispirando il progetto di un nuovo tipo di collisore elettroni-positroni, il collisore lineare produttore di mesoni *B*. Studi per tale macchina sono in corso all'Università della California a Los Angeles e in Italia. L'obiettivo è quello di produrre

Il decadimento del mesone *B* può produrre una violazione della parità di carica. Le collisioni protone-protone in un acceleratore producono coppie mesone-antimesone. Un mesone B^0 può decadere per mescolamento di sapori in un $\bar{B}^0$ (*a destra*), o viceversa (*a sinistra*). Tali particelle decadono poi in muoni, antimuoni, quark e neutrini. Se il tasso di decadimento in muoni (*a destra*) differisce dal tasso di decadimento in antimuoni (*a sinistra*), la *CP* è violata. (In questo caso la carica si conserva, come si vede sommando i diagrammi.)

più di un miliardo di quark e di antiquark bellezza all'anno.

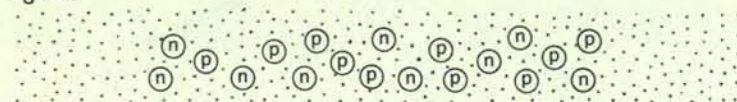
Il secondo importante mistero che potrebbe essere risolto chiamando in causa una quarta famiglia di particelle è l'origine della gerarchia delle masse delle particelle. La speranza è che la quarta famiglia si riveli un caso particolare e che le masse delle prime tre famiglie siano «generate» da interazioni con la quarta. Questa ipotesi, avanzata per la prima volta da Harald Fritzsch dell'Università di Monaco, correla la differenza di massa dei quark con una supposta relazione tra la massa dei quark e le loro transizioni. Tutti i dati sperimentali finora ottenuti sono in accordo con il modello di Fritzsch.

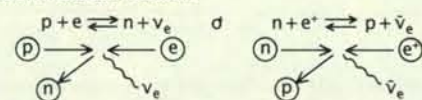
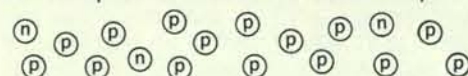
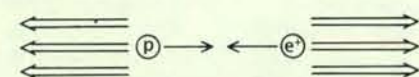
Fritzsch propone inoltre che il rapporto tra le masse dei due nuovi quark sia quattro. Questo valore è il rapporto tra i quadrati delle cariche dei quark - $(2/3)^2/(-1/3)^2$ - prevedibile nel caso in cui sia l'interazione elettromagnetica a dare origine alla massa dei quark.

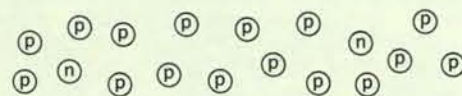
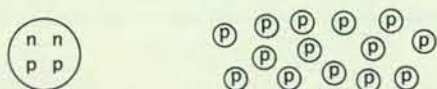
Alcuni fisici teorici ritengono che debba esistere un nuovo quark in prossimità

In base al meccanismo dell'altalena si può ipotizzare che la massa di ciascun neutrino (m_ν) sia correlata alla massa del leptone carico a esso associato (*m*) dalla formula $m_\nu = m^2 / \Omega$, dove Ω è una scala di massa sconosciuta, rappresentata nell'illustrazione dalla lunghezza del braccio dell'altalena. Per esempio, poiché si sa che la massa del neutrino elettronico è inferiore a 16 eV mentre quella dell'elettrone è 0,5 MeV, l'equazione dell'altalena impone che Ω abbia almeno il valore di 16 GeV. La massa della particella tau è 1,8 GeV e i limiti cosmologici impongono che la massa del neutrino tauonico sia minore di 65 eV. Inserendo nell'equazione dell'altalena tali valori, si trova per Ω il limite inferiore più rigoroso di 5 × 10⁷ GeV. Se Ω è correlata a un quarto leptone di grande massa, il meccanismo dell'altalena illustra come questa grande massa potrebbe generare le masse estremamente piccole dei neutrini.

LA PRODUZIONE DI ELIO NEL BIG BANG E LE FAMIGLIE DI NEUTRINI

1. Si supponga che l'universo sia composto solo da neutroni (*n*) e protoni (*p*), con quantità molto maggiori di elettroni (e^-), positroni (e^+), neutrini e antineutrini (ν_e, ν_μ, ν_τ, $\bar{\nu}_e$, $\bar{\nu}_\mu$, $\bar{\nu}_\tau$) e fotoni (γ), tutti indicati sotto da punti. A un'epoca molto anteriore a un secondo dopo il big bang e a temperature molto maggiori di 10^{10} kelvin, neutroni e protoni sono presenti in numero quasi uguale:

2. Neutroni e protoni sono continuamente trasformati l'uno nell'altro dalle cosiddette reazioni nucleari deboli:

$p + e^- \rightleftarrows n + \nu_e$ o $n + e^+ \rightleftarrows p + \bar{\nu}_e$

3. Dato che i neutroni sono leggermente più pesanti dei protoni, sono più difficili da produrre dal punto di vista energetico, perciò le trasformazioni *n-p* nella fase (2) danno origine a un lieve eccesso di protoni. Via via che l'universo si espande e si raffredda, è disponibile sempre meno energia per la produzione di neutroni, perciò le reazioni deboli accentuano sempre più l'eccesso di protoni. Circa un secondo dopo il big bang e a una temperatura intorno a 10^{10} kelvin, i protoni sono più abbondanti dei neutroni di cinque a uno:

4. A questo punto la velocità di espansione dell'universo supera la velocità sempre decrescente delle reazioni deboli, cosicché le collisioni tra particelle virtualmente cessano:

Non vi è più conversione di neutroni in protoni; il rapporto 1:5 è congelato.

5. I neutroni sono radioattivi e decadono in protoni. La vita media del neutrone è di circa 15 minuti, cosicché dopo circa tre minuti un terzo dei neutroni è decaduto in protoni, lasciando un *n* ogni otto *p*:

6. Tre minuti dopo il big bang la temperatura è scesa a 10^9 kelvin, abbastanza bassa perché il nucleo del deuterio (*n,p*) sia stabile. Il deuterio viene poi rapidamente trasformato in elio (*2n,2p*). Dato che la sintesi dell'elio richiede uguali quantità di protoni e di neutroni, essa termina quando non vi sono più neutroni disponibili:

Poiché neutroni e protoni hanno masse quasi uguali, circa 4/16, ovvero il 25 per cento, della massa dell'universo sono rappresentati da elio e il rimanente 75 per cento da protoni (nuclei di idrogeno).

7. Maggiore è il numero di famiglie di neutrini, più elevata è la velocità di espansione dell'universo. La fase (4) perciò avviene prima e a una temperatura più elevata quando sono presenti più neutroni; le fasi (5) e (6) hanno luogo in presenza di un maggior numero di neutroni e viene sintetizzato più elio. Le osservazioni astronomiche, tuttavia, limitano l'elio a meno del 25 per cento della massa totale dell'universo e questo, a sua volta, indica che non vi sono più di quattro famiglie di neutrini.

di 246 GeV. Uno degli aspetti notevoli del modello standard è la sua previsione che a energie abbastanza elevate le varie forze cominciano a unificarsi. In particolare, la forza elettromagnetica, la forza nucleare debole e la forza nucleare forte dovrebbero diventare una sola forza nell'ambito della «grande unificazione». Le forze dovrebbero unificarsi all'incredibile energia di 10^{15} GeV, notevolmente al di là di quanto si potrà mai ottenere con un acceleratore sulla Terra. L'estrapolazione dei valori misurati dei parametri fondamentali dalle basse energie alla scala energetica della grande unificazione richiederebbe l'esistenza di un nuovo quark pesante per dare coerenza alla teoria.

Si è visto inoltre che i valori misurati della massa delle particelle W e Z^0 possono fornire limiti alla differenza di massa tra i due membri della quarta famiglia di quark. I dati attuali indicano che la differenza di massa tra i due quark della quarta famiglia dovrebbe essere inferiore a 180 GeV. Perciò se esiste un membro della famiglia in prossimità di 246 GeV, la massa dell'altro membro dovrebbe trovarsi all'incirca o a 426 GeV o a 66 GeV, e quest'ultimo valore si trova ben all'interno dell'intervallo di energia che gli acceleratori saranno in grado di esplorare nel prossimo futuro.

Le masse dei neutrini

Se esiste la quarta famiglia di quark, come saranno i corrispondenti leptoni? Qui l'interesse si focalizza sulle masse dei neutrini. I limiti cosmologici citati in precedenza richiedono che i neutrini abbiano una massa inferiore a 65 eV, un intervallo che comprende anche il valore zero. Se si confronta tale valore con la massa della particella W, si ha un rapporto di massa pari a un miliardo. Come si può spiegare la massa incredibilmente piccola del neutrino? Esistono due diversi punti di vista: o la massa del neutrino è esattamente nulla a causa di qualche principio fondamentale non ancora scoperto, oppure la piccola massa del neutrino è la conseguenza di un'altra massa estremamente grande.

Quest'ultimo punto di vista si basa su quello che è stato chiamato «meccanismo dell'altalena», proposto da Murray Gell-Mann del California Institute of Technology, Pierre M. Ramond dell'Università della Florida e Richard C. Slansky del Los Alamos National Laboratory. L'inconveniente del meccanismo dell'altalena è quello di essere costruito *ad hoc*, mentre il suo vantaggio è quello di essere estremamente semplice. Esso ipotizza che la massa del neutrino elettronico sia uguale al rapporto tra il quadrato della massa dell'elettrone e una grande massa sconosciuta. La massa dell'elettrone è fissa, perciò più grande sarà la massa sconosciuta, minore risulterà la massa del neutrino; da questo fatto deriva il nome di meccanismo dell'altalena (*si veda l'illustrazione a pagina 33 in basso*).

Per fare un esempio, i limiti cosmologici ricavati dallo studio delle supernove impongono che la massa del neutrino elettronico sia minore di 16 eV. Il quadrato della massa dell'elettrone è di circa 250 GeV^2. La soluzione dell'equazione dell'altalena richiede quindi che la massa sconosciuta sia maggiore di 16 GeV. Quando si eseguono i conti con la massa di 1,8 GeV della particella tau e con il limite superiore cosmologico di 65 eV per la massa del neutrino tauonico, si trova un limite inferiore più rigoroso di 5×10^7 GeV.

A causa del meccanismo dell'altalena la massa incredibilmente minuscola dei neutrini sembra essere la conseguenza di una scala di massa molto maggiore di ciò che si può ottenere con gli attuali acceleratori a collisione di fasci. Se il meccanismo è corretto, queste scale di massa dovrebbero essere associate a nuove particelle, forse un quarto quark o un quarto leptone. Il meccanismo dell'altalena dovrebbe avere in questo caso conseguenze cosmologiche: esso fa sorgere la possibilità che sia il quarto neutrino a fornire la cosiddetta massa mancante necessaria per «chiudere» l'universo.

L'orientamento teorico attuale richiede che la densità di massa dell'universo sia appena sufficiente per arrestare e quindi invertire l'attuale espansione, nel qual caso l'universo si dice «chiuso». I dati disponibili indicano tuttavia che la densità di massa osservata nell'universo è compresa tra il 10 e il 20 per cento soltanto di tale valore critico. Gli astronomi sono perciò attualmente impegnati in una ricerca di vasto respiro della «massa mancante».

Un neutrino che fornisca la massa mancante non può essere troppo pesante. I neutrini sono persino più abbondanti dei fotoni - ve ne sono parecchi miliardi per ogni protone, elettrone e neutrone - e se un qualsiasi tipo di neutrino avesse una massa uguale a 65 eV, questo basterebbe da solo a chiudere l'universo. Se la massa del neutrino fosse molto superiore a tale valore, l'attrazione gravitazionale risultante sarebbe sufficiente a rallentare notevolmente la velocità di espansione dell'universo. Il fatto che non sia stato osservato alcun effetto del genere ha indotto la maggior parte dei fisici ad accettare come limite superiore il valore di 65 eV.

Si sa dai dati sperimentali che un quarto leptone carico dovrebbe avere una massa superiore a 41 GeV. Assumendo tale valore per la massa del leptone carico e 65 eV per la massa del quarto neutrino, il meccanismo dell'altalena fornisce un valore di $2,5 \times 10^{10}$ GeV per la scala di massa sconosciuta. Supponendo che questa singola scala di massa generi le masse di tutti i neutrini, si può in questo caso calcolare con il meccanismo dell'altalena che la massa del neutrino elettronico, quella del neutrino muonico e quella del neutrino tauonico devono essere inferiori a 10^{-8} eV, 4×10^{-4} eV e 0,1 eV rispettivamente. Se questo ragionamento è corretto, il quarto neutrino potrebbe fornire la massa mancante, ma i tre neutrini conosciuti sarebbero di gran lunga troppo leggeri per avere un qualsiasi effetto.

Gli esperimenti in corso

Le argomentazioni appena citate sull'esistenza di una quarta famiglia di quark e di leptoni sono chiaramente speculative; tuttavia sono in corso alcune ricerche dirette. Una tecnica suggerita per la prima volta da me e da Carlo Rubbia del CERN ricorre al decadimento delle particelle W per determinare la massa di un possibile quarto leptone carico o per porvi un limite. Recenti esperimenti condotti al CERN forniscono un limite inferiore di 41 GeV. Si noti che tale valore è da uno a due ordini di grandezza superiore alla massa della particella tau, come ci si attende dalla gerarchia delle masse discussa in precedenza. Se la massa di un ulteriore quark o di un leptone carico è inferiore a 70 GeV, le macchine attuali potrebbero forse rivelarli in un prossimo futuro. Altrimenti i fisici dovranno attendere il supercollisore a magneti superconduttori (Superconducting Supercollider) o il Large Hadron Collider la cui costruzione è in programma al CERN.

Ci si potrebbe anche chiedere se, una volta scoperta la quarta famiglia di quark e leptoni, vi sarà la prospettiva di trovarne una quinta. È questa la domanda che si pongono i cosmologi e i fisici delle particelle nei tentativi in atto di contare le famiglie di neutrini.

Dal punto di vista cosmologico, il numero di famiglie di neutrini ha un effetto profondo sulla produzione di isotopi leggeri nel processo di nucleosintesi primordiale che avvenne nei primissimi istanti dopo il big bang. Le abbondanze attuali di questi isotopi, in particolare del deuterio e degli isotopi dell'elio, dipendono dalla velocità di espansione dell'universo rispetto alla velocità delle reazioni nucleari che producevano gli isotopi stessi. La velocità di espansione dell'universo dipende a sua volta dal numero di specie di particelle esistenti, tra cui le famiglie di neutrini. Maggiore è il numero di famiglie di neutrini esistenti, più rapida è l'espansione dell'universo e più elio viene prodotto. Il confronto tra la quantità di elio prodotto secondo i calcoli sulla nucleosintesi primordiale e i limiti superiori osservati pone un vincolo al numero delle possibili famiglie di neutrini (*si veda la «finestra» nella pagina a fronte*). Si noti che queste considerazioni limitano il numero di famiglie di neutrini a quattro o al massimo a cinque. Se si suppone che l'accoppiamento quark-leptone del modello standard continui nelle famiglie successive, ne consegue che il numero di famiglie di quark è limitato anch'esso a quattro o cinque.

Tali limiti cosmologici sulle specie di particelle sono stati tradizionalmente visti con molta prudenza dai fisici delle particelle. Ora però gli esperimenti di laboratorio stanno giungendo alla stessa conclusione. Questi esperimenti utilizzano la particella Z^0, che può decadere in tutte le famiglie di neutrini esistenti; più famiglie di neutrini vi sono, più rapido è il decadimento. Perciò, misurando la vita media della particella Z^0, è possibile determinare il numero di famiglie di neutrini esistenti in natura. I risultati preliminari che sono stati ottenuti al CERN e presso gli acceleratori SLAC e DESY hanno già limitato a cinque il numero di famiglie di neutrini. Ulteriori miglioramenti finiranno senza dubbio per fornire un numero esatto e non solamente un limite superiore: la prova diretta pro o contro una quarta famiglia di quark e di leptoni.

Se si scoprirà una nuova famiglia oltre verità e bellezza, i fisici si chiederanno senza dubbio: perché debbono esservene quattro? Perché le masse delle particelle della quarta famiglia sono così grandi? Perché quattro (un numero pari) è anche il numero di dimensioni dello spazio-tempo? È forse questo un risultato della teoria delle supercorde, che precisa il numero delle possibili dimensioni? Certamente queste domande possono continuare all'infinito. Tuttavia la sola cosa che gli scienziati possono fare è quella di ridurre a pochi un gran numero di problemi. Se l'introduzione della quarta famiglia di quark e di leptoni sarà in grado di spiegare la distribuzione di massa all'interno delle prime tre famiglie di particelle, le probabilità di transizione quark-quark, la massa mancante e la natura della violazione della parità di carica, allora sarà stato compiuto un grande passo in avanti.

BIBLIOGRAFIA

FRITZSCH H., *Weak-Interaction Mixing in the Six-Quark Theory* in «Physics Letters», 73B, n. 3, 27 febbraio 1978.

CLINE D. B. e RUBBIA C., *Search for High-Mass Sequential Leptons through the Change of Charge Asymmetry in $W^\pm \leftarrow l^\pm$ + (Neutrinos) Decay* in «Physics Letters», 127B, n. 3, 4, 28 luglio 1983.

YANG J., TURNER M. S., STEIGMAN G., SCHRAMM D. N. e OLIVE K. A., *Primordial Nucleosynthesis: A Critical Comparison of Theory and Observation* in «The Astrophysical Journal», 281, n. 2, 15 giugno 1984.

The First International Symposium on the Fourth Family of Quarks and Leptons in «Annals of the New York Academy of Sciences», 518, 1988.

Apprendimento e memoria *in vitro*

La ricostruzione in coltura del circuito nervoso alla base di un semplice comportamento del gasteropodo marino Aplysia californica *aiuta a chiarire i meccanismi molecolari della memoria e dell'apprendimento*

di Pier Giorgio Montarolo e Samuel Schacher

«Per quanto meraviglioso sia, il sistema nervoso di un invertebrato, come la sanguisuga, l'*Aplysia*, il polpo, non arriva a essere un modello adeguato del cervello umano, con le sue capacità concettuali, conoscitive, volitive e autocoscienti. Pur tuttavia, le indagini a questi livelli sono di importanza fondamentale, in quanto la conoscenza delle unità costitutive e delle loro semplici interconnessioni contribuisce a farci comprendere come funzionano i sistemi nervosi più complessi.» (John C. Eccles)

I mezzi che l'uomo, come la maggior parte delle specie animali, ha a disposizione per adattarsi all'ambiente sono essenzialmente due: l'evoluzione biologica e l'apprendimento. L'evoluzione biologica, legata al corredo cromosomico, è un processo lento, misurabile in termini di migliaia se non di milioni di anni negli organismi superiori; l'apprendimento è, invece, un processo rapido e il suo ambito temporale è l'arco di vita dell'organismo. Per poter sopravvivere un animale deve essere in grado di riconoscere certe relazioni chiave tra situazioni ed eventi esterni. Deve, per esempio, saper distinguere la preda dal predatore, un frutto buono da uno velenoso. Vi sono due possibilità per arrivare a queste conoscenze. La capacità di distinguere tra due alternative può essere programmata nel sistema nervoso centrale fin dalla nascita, oppure si può acquisire tramite l'apprendimento.

Gli animali molto semplici, come per esempio i nematodi, dotati di un sistema nervoso elementare, hanno un repertorio comportamentale fisso, mentre gli organismi più complessi hanno la capacità di adattarsi alle nuove situazioni che, di volta in volta, si presentano nel corso della loro esistenza. Questa capacità di adattamento deriva dal fatto che l'esperienza può modificare il sistema nervoso e quindi il comportamento dell'animale. In questo concetto di adattabilità è racchiuso il significato dei termini apprendimento e memoria. Utilizzando le parole di Larry Squire dell'Università della California, per apprendimento si intende il processo attraverso il quale un organismo acquisisce nuove informazioni; la memoria è l'immagazzinamento di queste informazioni nel sistema nervoso centrale in modo che possano essere successivamente utilizzate. In tale modo l'animale trae vantaggio dall'esperienza e il suo comportamento futuro risulta più adatto all'ambiente in cui esso vive.

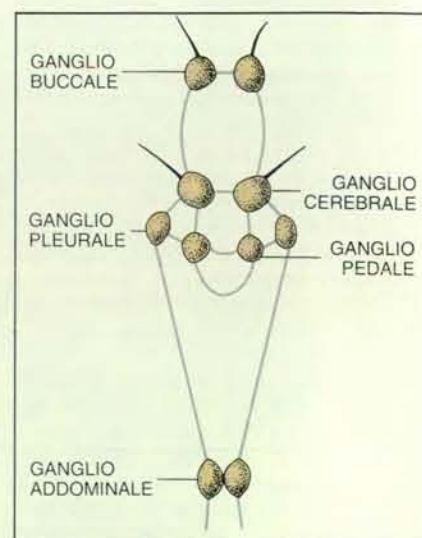
Alla fine del secolo scorso lo psicologo tedesco Hermann Ebbinghaus affrontò per primo lo studio dell'apprendimento e della memoria usando la metodologia sperimentale. Egli sviluppò alcuni metodi per misurare la capacità di memorizzazione dell'uomo. Basandosi sugli esperimenti di Ebbinghaus e sui propri studi, lo psicologo americano William James intuì che il processo mnemonico doveva essere costituito da due componenti distinte: una a breve termine e una a lungo termine (egli le chiamò rispettivamente memoria primaria e memoria secondaria). L'ipotesi di James fu confermata molto presto da altri due psicologi tedeschi, Georg Elias Müller e Alfons Pilzecher. Essi osservarono che una informazione, per poter essere ricordata sufficientemente a lungo, ha bisogno di un periodo di consolidamento durante il quale il ricordo è labile e l'informazione può essere facilmente persa. Una volta che il consolidamento è avvenuto (si può consolidare il ricordo con la presentazione iterata dell'informazione) il ricordo è pressoché incancellabile.

Osservazioni cliniche contribuirono a sostenere l'ipotesi che la memoria consista di una componente labile e di una componente permanente. Era già noto che alcuni traumi cranici producono amnesia retrograda. Una contusione cerebrale può provocare un'amnesia che riguarda solo gli avvenimenti che hanno preceduto di poco (nell'ordine di ore) l'episodio traumatico, mentre non viene intaccato il ricordo degli avvenimenti risalenti a giorni e ad anni prima del trauma. Anche i pazienti epilettici spesso non ricordano ciò che è accaduto prima dell'attacco convulsivo.

Il concetto di due processi mnemonici distinti, uno più labile e uno più stabile, oppure - come più correntemente si dice - di una memoria a breve termine e di una memoria a lungo termine, restava comunque una pura osservazione fenomenologica; non ci si addentrava nell'analisi dei meccanismi che ne sono alla base. La modificazione del comportamento è soltanto una delle manifestazioni dell'apprendimento e, per studiare i cambiamenti che avvengono all'interno del sistema nervoso centrale, è necessario ricorrere a un metodo sperimentale diverso che non comporti la sola analisi comportamentale.

Una possibile strategia per comprendere i meccanismi cellulari dell'apprendimento e della memoria consiste nel descrivere esattamente il circuito nervoso che presiede a un comportamento plastico, vale a dire modificabile dall'esperienza. Successivamente si deve analizzare la natura delle modificazioni indotte dall'apprendimento nella catena neuronale. Questa non è un'impresa facile se si studiano i fenomeni nei vertebrati superiori. Infatti il numero di neuroni coinvolti in comportamenti semplici, come certi riflessi di difesa dei mammiferi, è dell'ordine di diverse migliaia. Il numero sale a ordini di grandezza superiori se si analizzano comportamenti più complessi. Vi sono comunque ricercatori come Masao Ito dell'Università di Tokyo e Richard Thompson della Stanford University, che non si sono lasciati scoraggiare da queste difficoltà e hanno descritto, in alcuni mammiferi, circuiti neuronali coinvolti nei processi di apprendimento.

È possibile superare in parte queste difficoltà individuando comportamenti plastici in animali dotati di sistema nervoso molto semplice. Questo è il principio da cui alla fine degli anni sessanta partì Eric R. Kandel della Columbia University. Come oggetto di studio egli scelse *Aplysia californica*, un mollusco gasteropodo marino che ha un sistema nervoso centrale relativamente semplice formato da circa 20 000 neuroni, distribuiti principalmente in nove raggruppamenti, i gangli. I corpi cellulari di alcuni neuroni di *Aplysia* possono raggiungere il diametro di 1,5 millimetri (le più grosse cellule nervose dell'uomo hanno

Nel sistema nervoso centrale di *Aplysia*, qui schematizzato, le quattro coppie di gangli - buccali, cerebrali, pleurali e pedali - costituiscono l'anello esofageo. Generalmente si ritiene che questo gruppo di gangli localizzati nella testa controlli la maggior parte delle funzioni somatiche (per esempio la locomozione e l'assunzione di cibo) mentre le funzioni viscerali (riproduzione, respirazione, circolazione ed escrezione) sono controllate dal ganglio addominale. Quest'ultimo, che risulta dalla fusione di molti piccoli agglomerati neuronali, presiede anche al movimento (a scopo difensivo) degli organi alloggiati nella cavità del mantello e controlla pure il processo di deposizione delle uova, il bilancio idrico e l'emissione dell'inchiostro. Si valuta che il sistema nervoso di *Aplysia* sia formato da circa 20 000 neuroni. Il cervello umano ne contiene un numero maggiore di otto ordini di grandezza.

La lunghezza di questo esemplare giovanile di *Aplysia californica*, allevato nel Marine Biological Laboratory di Woods Hole e qui presentato in una veduta dorsale, è di circa due centimetri per pochi grammi di peso. Se si eccettuano le dimensioni, l'esemplare giovanile è indistinguibile dall'animale adulto che può raggiungere il metro di lunghezza e un peso di sette chilogrammi. I parapodi distesi permettono di vedere la cavità del mantello dove alloggia l'organo respiratorio, la branchia. In questa fotografia, la branchia appare ricoperta dalla plica del mantello. Il mantello, posteriormente, forma una estroflessione imbutiforme, il sifone, attraverso la quale avviene l'espulsione dell'acqua marina e delle feci. Alcune specie del genere *Aplysia* sono presenti anche nel Mediterraneo.

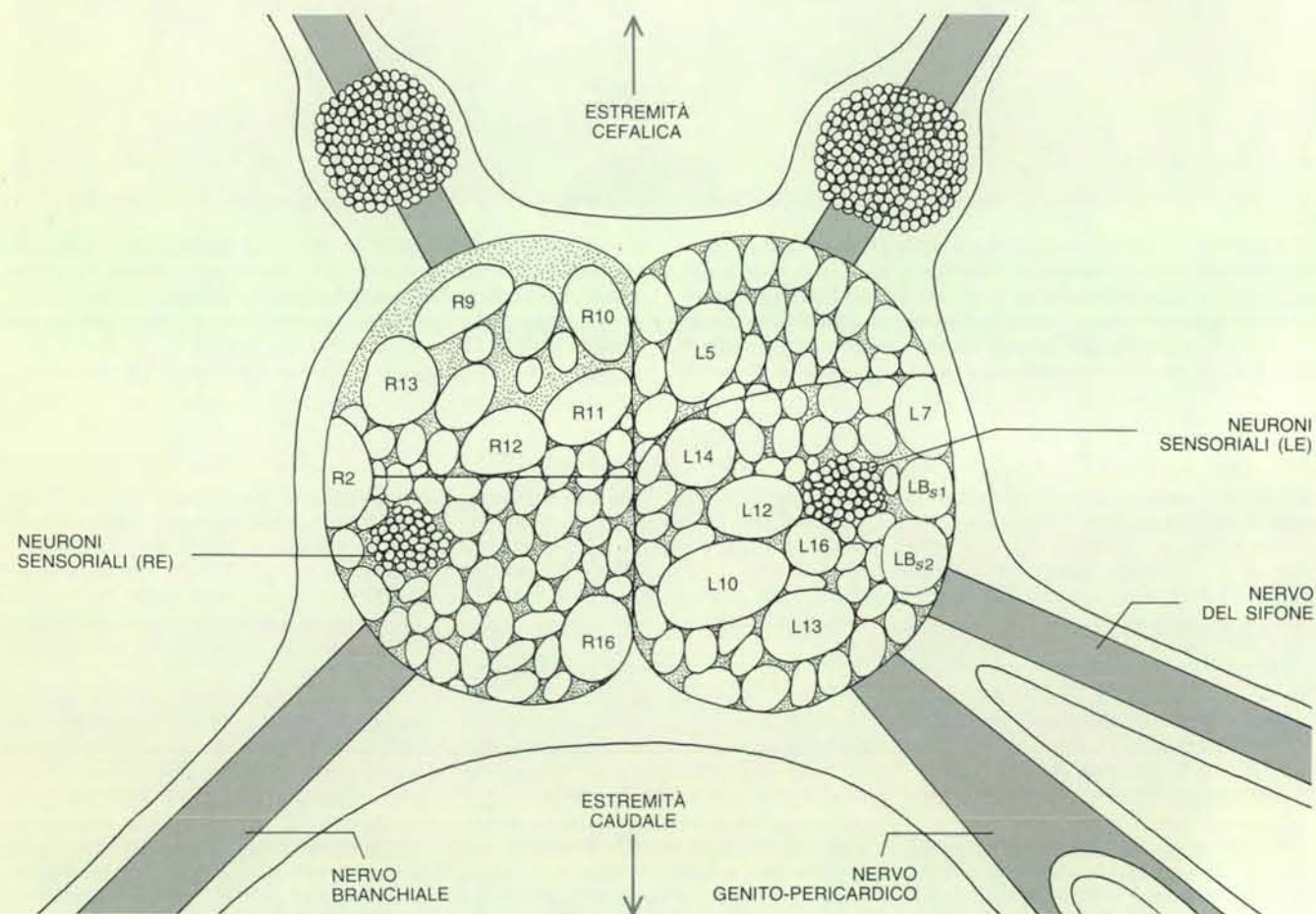
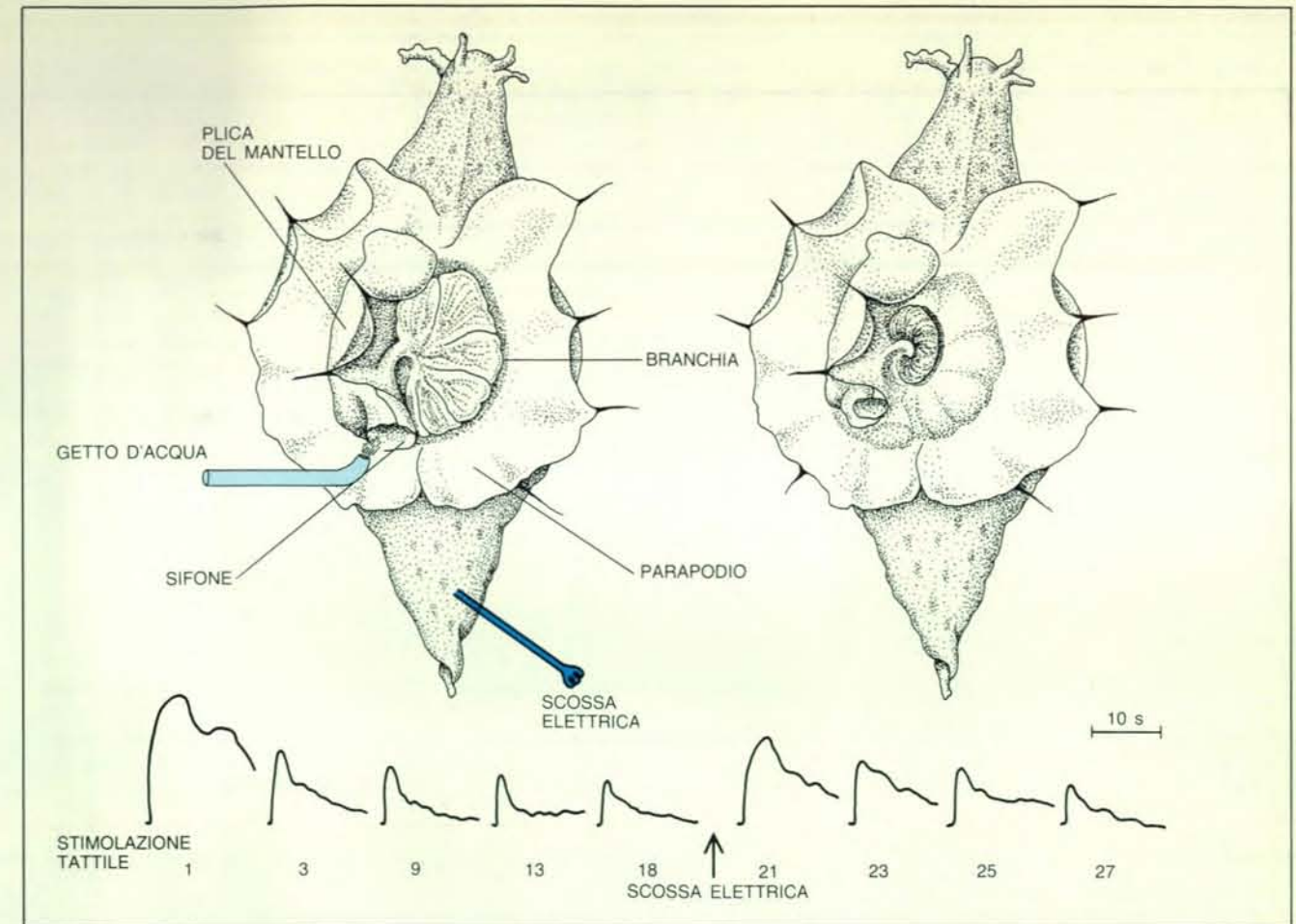
un diametro circa 30 volte inferiore).

La scoperta che questo semplice organismo, così distante dai mammiferi nella scala zoologica, mostrasse alcune forme di apprendimento e di memoria non fu un evento inaspettato. L'apprendimento, grazie allo studio di etologi come Konrad Lorenz e Nikolaas Timbergen, si è rivelato una proprietà molto diffusa nel regno animale. Ciò non significa che i processi di apprendimento nell'uomo e in *Aplysia* siano uguali, ma sta a indicare che, nei due organismi, si possono riconoscere alcuni meccanismi comuni, conservatisi attraverso la filogenesi. Sono proprio questi aspetti comuni che si vuole identificare e descrivere, utilizzando animali dotati di un sistema nervoso centrale molto più semplice di quello dell'uomo.

Il comportamento analizzato da Kandel e dai suoi collaboratori è un semplice riflesso di difesa: il riflesso di retrazione della branchia e del sifone. La branchia è l'organo respiratorio; il sifone è un'estroflessione imbutiforme di una struttura chiamata mantello, la cui parte membranosa ricopre la branchia. Una stimolazione tattile del sifone produce una rapida e prolungata retrazione della branchia e del sifone stesso all'interno della cavità del mantello. Circa 40 neuroni entrano nella costituzione di questo arco riflesso. Un semplice comportamento di questo tipo risulta modificabile da forme di apprendimento associativo e non associativo; le più studiate sono due forme non associative: l'assuefazione e la sensibilizzazione.

L'assuefazione consiste nella progressiva riduzione di una risposta comportamentale in seguito alla presentazione ripetuta di uno stimolo innocuo. Questo processo di apprendimento è molto diffuso nel regno animale ed è il primo a comparire nel bambino. Per suo tramite l'animale impara a ignorare gli stimoli privi di significato per la sopravvivenza.

Una decina di stimolazioni tattili del sifone, ripetute a intervalli regolari di 20-30 secondi, produce una progressiva riduzione della durata e dell'ampiezza della retrazione fino alla soppressione del riflesso. La diminuzione o la scomparsa della risposta può persistere per poche ore o per alcuni giorni a seconda della durata del protocollo di stimolazione utilizzato. Si parla di assuefazione rispettivamente a breve e a lungo termine.

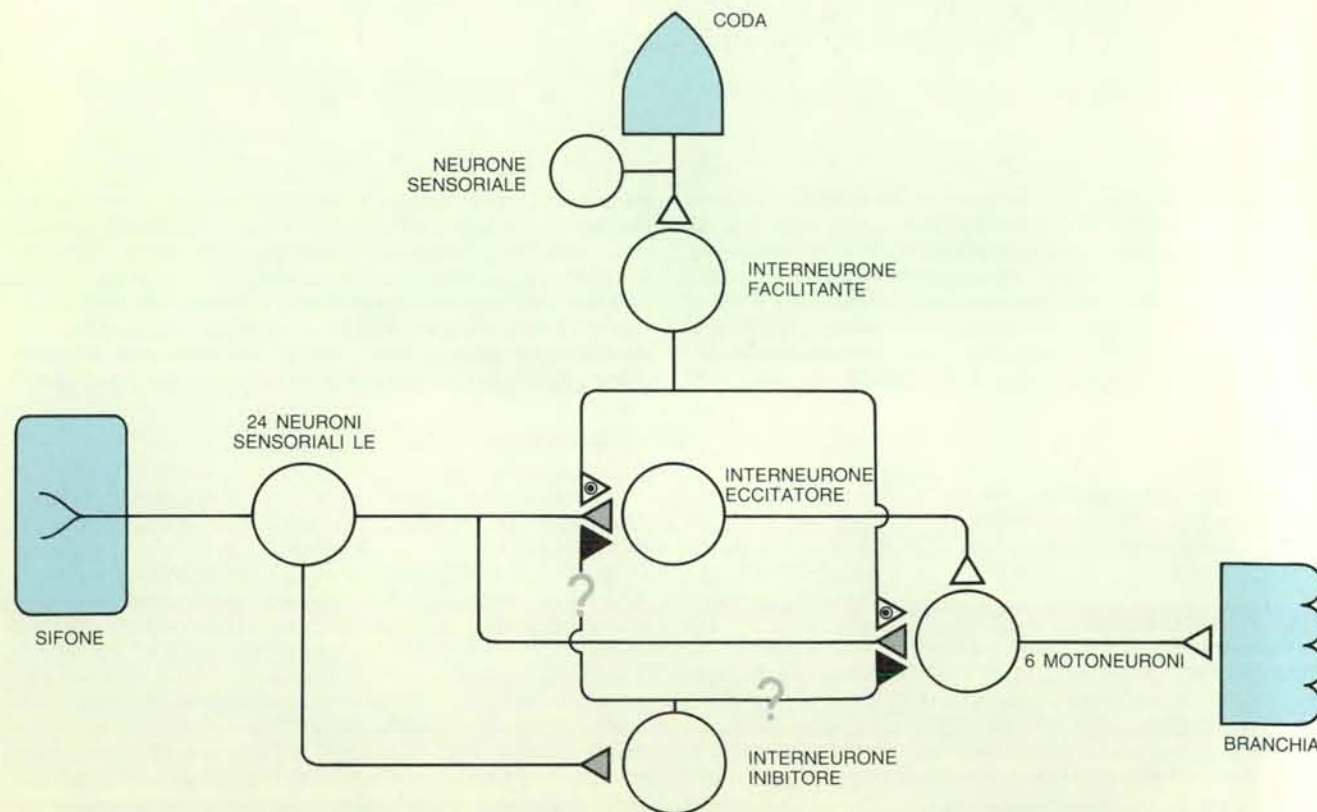
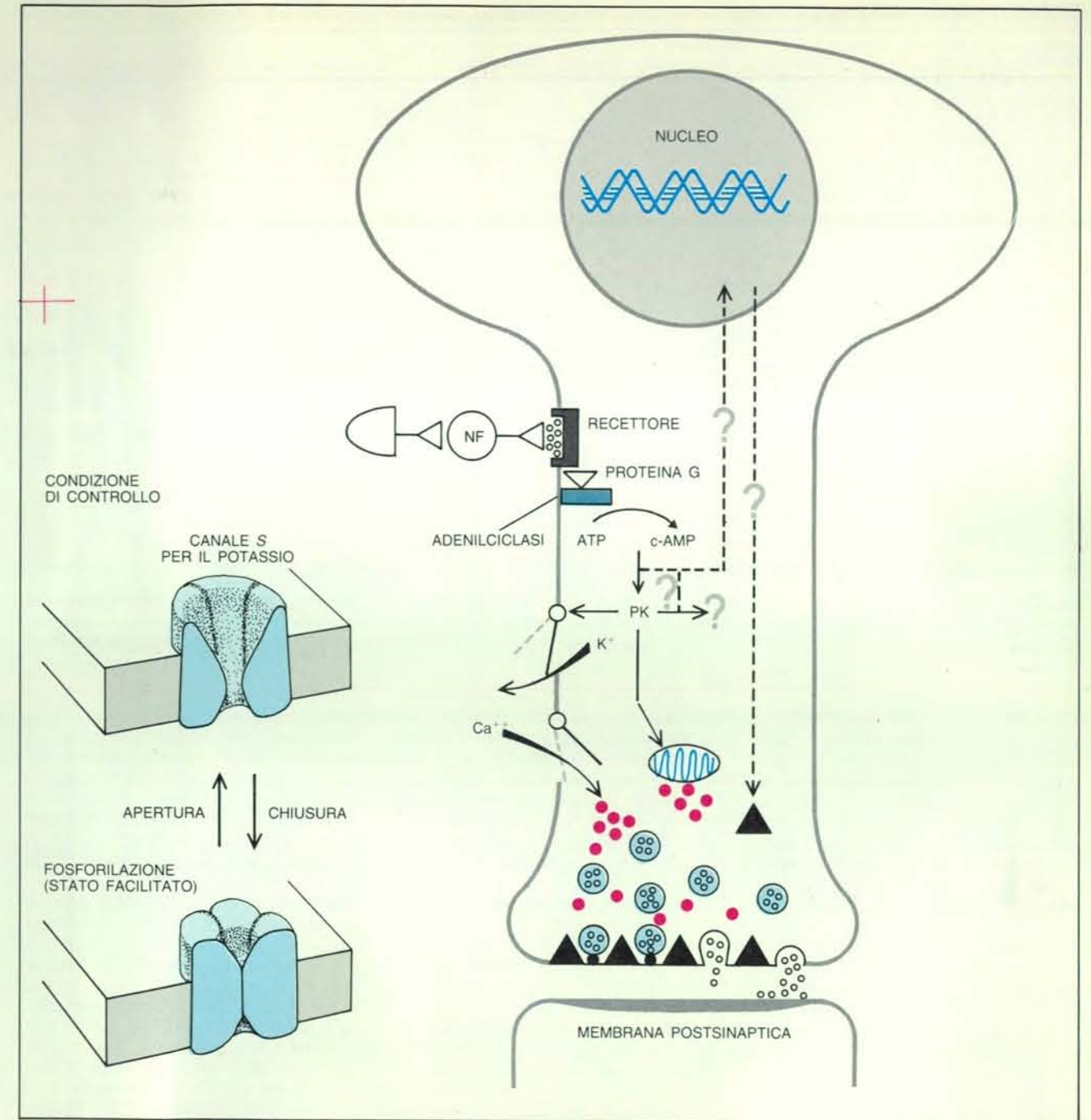
Dapprima Kandel e i suoi collaboratori hanno studiato i meccanismi cellulari alla base dell'assuefazione a breve termine e hanno identificato la causa della riduzione della risposta comportamentale nella diminuzione della trasmissione a livello delle sinapsi - i punti specializzati di contatto tra due cellule nervose - che i neuroni sensoriali del gruppo LE (L sta per *left*, sinistra) formano con le cellule bersaglio: gli interneuroni e i motoneuroni. In seguito Mark Klein della Columbia University ha scoperto che la riduzione della trasmissione sinaptica era dovuta alla progressiva diminuzione dell'entrata di ioni calcio (Ca^{++}) nella cellula sensoriale durante i potenziali d'azione generati dalle ripetute stimolazioni tattili. La riduzione del flusso di Ca^{++} diminuisce la quantità di neurotrasmettitore liberato dalle terminazioni dei neuroni sensoriali.

Il ganglio addominale, di cui è mostrato il lato ventrale, è suddiviso in due emigangli e le cellule identificate sono indicate con L (da *left*, sinistra) o R (da *right*, destra) a seconda che si trovino nell'emiganglio sinistro o destro. La presenza di una seconda lettera indica un raggruppamento di cellule con funzioni analoghe; per esempio LB, LE, mentre il deponente indica la funzione. Così le sigle LB$_{s1}$ e LB$_{s2}$ indicano che le cellule appartenenti allo stesso raggruppamento B sono motoneuroni del sifone (*s*). Fra questi ultimi il più facile da individuare per costanza di posizione, pigmentazione e forma è quello siglato con L7. Le cellule sensoriali usate nella ricostruzione del circuito *in vitro* appartengono al raggruppamento LE. Questi neuroni rispondono a stimoli tattili e quindi sono meccanocettori.

È illustrato come avvengono assuefazione e sensibilizzazione a breve termine del riflesso di retrazione della branchia in *Aplysia*. Si può suscitare la risposta riflessa di retrazione della branchia e del sifone con una stimolazione tattile effettuata con un getto d'acqua diretto sulla cute del sifone. Come risposta, la branchia e il sifone si retraggono nella cavità del mantello (*in alto a destra*). Si può misurare l'entità e la durata della retrazione mediante una fotocellula posta sotto la branchia e tramite un opportuno circuito elettronico si ottiene un segnale in volt (*in basso*) che è proporzionale alla durata e all'ampiezza della contrazione. Stimolando ripetutamente una volta al minuto, si assiste a una progressiva riduzione di questa risposta difensiva (assuefazione). Una scossa elettrica alla coda fa sì che il successivo getto d'acqua produca una contrazione vigorosa e duratura paragonabile, e a volte più intensa, alla prima (sensibilizzazione o disassuefazione). Continuando con lo stesso ritmo le stimolazioni, l'entità della risposta riflessa torna a ridursi.

Dai classici esperimenti, condotti negli anni cinquanta da Bernard Katz dell'University College di Londra si sono accumulati molti dati sperimentali che hanno chiarito come avviene la trasmissione del segnale a livello della sinapsi chimica. Quando il potenziale d'azione raggiunge la terminazione del neurone presinaptico, esso produce la secrezione di una piccola quantità di un neurotrasmettitore o mediatore chimico. Il potenziale d'azione consiste in una rapidissima e brevissima inversione del potenziale elettrico intracellulare che, da un valore compreso tra −70 e −90 millivolt (mV) in condizione di riposo, raggiunge un valore di circa +30 millivolt. Questa inversione di polarità è prodotta principalmente dall'entrata di ioni sodio (Na^+) e, in misura minore, di ioni calcio (Ca^{++}). Il potenziale elettrico ritorna ai valori di riposo grazie all'uscita dalla cellula di ioni potassio (K^+). Il neurotrasmettitore non viene liberato come singola molecola, bensì in pacchetti multimolecolari chiamati quanti. Si ritiene che i quanti siano contenuti in strutture subcellulari, le vescicole, presenti in grande quantità nelle terminazioni presinaptiche. Le vescicole si aprono e liberano il loro contenuto nello spazio intersinaptico solo se vengono a contatto con strutture specializzate della parte interna della membrana cellulare, i siti attivi (o zone attive). Il quanto di trasmettitore, interagendo con lo specifico recettore sulla membrana del neurone postsinaptico, causa una piccola variazione (dell'ordine di microvolt) del potenziale di membrana: il potenziale postsinaptico in miniatura. La contemporanea liberazione di migliaia di quanti produce, invece, una più ampia variazione (dell'ordine di millivolt), il normale potenziale postsinaptico. Questa liberazione ingente di mediatore chimico è causata prevalentemente dall'ingresso, nella terminazione presinaptica, di Ca^{++} durante il potenziale d'azione. Il Ca^{++} infatti permette l'interazione delle vescicole con la zona attiva. Si può quindi capire come variazioni nell'ingresso di questo ione modulino la liberazione del neurotrasmettitore e, quindi, l'efficacia della trasmissione sinaptica.

Un'altra forma di plasticità comportamentale molto studiata dal gruppo della Columbia University è la sensibilizzazione. Questa modalità di apprendimento consiste nel potenziamento di una risposta a uno stimolo innocuo qualora l'animale abbia ricevuto in precedenza uno stimolo intenso o nocivo. L'animale impara a prestare attenzione a uno stimolo

innocuo in quanto esso può essere accompagnato da situazioni pericolose. Una scossa elettrica alla testa o alla coda di *Aplysia* fa sì che la successiva stimolazione tattile del sifone produca una contrazione della branchia più intensa e duratura di quella che si osserva in seguito a una analoga stimolazione prima della scossa. Il meccanismo cellulare di questo aumento della risposta comportamentale è in un certo qual modo opposto a quello responsabile dell'assuefazione. Infatti, la trasmissione del segnale a livello della sinapsi sensomotoria diminuisce nell'assuefazione, mentre viene potenziata nella sensibilizzazione. L'aumento è dovuto all'azione di neuroni facilitanti (NF), estrinseci al circuito del riflesso di retrazione della branchia e attivati dallo stimolo nocivo applicato alla testa o alla coda. Il fenomeno viene chiamato «facilitazione eterosinaptica». I neuroni facilitanti prendono contatto sinaptico con i meccanocettori dell'arco afferente del riflesso della branchia. Uno di questi neuroni è stato identificato e siglato L29.

Si hanno valide ragioni per ritenere che uno dei neuromediatori liberati dai neuroni facilitanti sia la serotonina (denominata anche 5-idrossitriptammina). L'azione della serotonina sulla cellula sensoriale è complessa. In base a esperimenti di elettrofisiologia e di biochimica condotti dai ricercatori della Columbia University, Kandel e James Schwartz hanno proposto il seguente modello per spiegare la sensibilizzazione a breve termine. La serotonina liberata dai neuroni facilitanti si lega a specifici recettori di membrana del neurone sensoriale del gruppo LE. Ne consegue un aumento del contenuto intracellulare di adenosinmonofosfato ciclico (c-AMP) che attiva la proteinchinasi AMP ciclico-dipendente, un enzima che catalizza la reazione di fosforilazione di diverse classi di proteine intracellulari. L'aggiunta di un gruppo fosforico a una proteina può modificarne la forma e attivarne la funzione. Una conseguenza diretta o indiretta della fosforilazione indotta dalla serotonina è il cambiamento conformazionale di una proteina di membrana che forma un poro nella membrana (canale *S*, da serotonina), permettendo il passaggio degli ioni K^+. In altre parole, il canale si chiude facendo diminuire l'uscita di questi ioni, con conseguente rallentamento della ripolarizzazione della membrana cellulare e incremento di durata del potenziale d'azione. In tale condizione, i canali per il calcio restano aperti più a lungo e si verifica, quindi, un maggior flusso di ioni Ca^{++} all'interno della terminazione presinaptica. Questo fa aumentare la liberazione del neurotrasmettitore da parte del neurone sensoriale. Ne risulta un potenziamento della trasmissione del segnale a livello delle sinapsi che i neuroni sensoriali formano con le cellule bersaglio.

Sostanzialmente si può osservare come, alla base di queste semplici forme di apprendimento e di memoria a breve termine, vi sia un processo di modulazione della trasmissione sinaptica. L'idea che l'apprendimento possa implicare modificazioni nelle connessioni tra le cellule di una rete neuronale non è nuova. I primi a suggerire questa possibilità furono, alla fine del secolo scorso, il neurologo italiano Eugenio Tanzi, il grande neurobiologo Santiago Ramón y Cajal e

Il circuito, qui semplificato, del riflesso di retrazione della branchia in *Aplysia* è formato dai 24 neuroni sensoriali del gruppo LE che formano una connessione monosinaptica con sei motoneuroni che innervano la muscolatura della branchia. Le cellule sensoriali prendono pure contatto sinaptico diretto con interneuroni eccitatori (se ne conoscono due) che, a loro volta, formano sinapsi con i motoneuroni. È stato identificato anche un interneurone inibitore la cui funzione non è stata ancora completamente chiarita. La sinapsi che il neurone sensoriale forma con le sue cellule bersaglio è il sito chiave la cui plasticità spiega le modificazioni comportamentali a cui il riflesso va incontro (assuefazione e sensibilizzazione). Le ripetute stimolazioni tattili del sifone, reclutando i meccanocettori LE, producono una progressiva riduzione della trasmissione sinaptica (depressione omosinaptica). La stimolazione elettrica della coda, tramite altri neuroni sensoriali, attiva alcuni interneuroni facilitanti che, formando sinapsi con i neuroni LE, potenziano la trasmissione sempre allo stesso livello al quale i neuroni LE agiscono con le loro cellule bersaglio (facilitazione eterosinaptica).

Alla base della facilitazione a breve e a lungo termine vi è una serie di eventi molecolari. La serotonina, o comunque altri neuromodulatori liberati dai neuroni facilitanti (NF) in seguito alla scossa elettrica alla coda o in qualunque altra parte del corpo di *Aplysia*, interagisce con specifici recettori di membrana del neurone sensoriale del gruppo LE. Tramite la modificazione conformazionale di una proteina di membrana, la proteina G, viene attivato un enzima, l'adenilciclasi, che catalizza la sintesi di adenosinmonofosfato ciclico (c-AMP) a partire da adenosintrifosfato (ATP). L'aumento del contenuto citosolico di AMP ciclico attiva l'enzima proteinchinasi AMP ciclico-dipendente, PK, che aggiunge un residuo fosforico a una grande varietà di proteine e, di conseguenza, esercita effetti molteplici a seconda delle funzioni delle proteine fosforilate. La fosforilazione causa una modificazione covalente della struttura proteica. Uno degli effetti, diretto o indiretto, è la chiusura del poro della proteina canale *S* (da serotonina) per il K^+ (*a sinistra*). La depolarizzazione legata al potenziale d'azione dura quindi più a lungo e il canale per il Ca^{++} (che è mantenuto aperto dalla depolarizzazione) resta aperto per un tempo più lungo, permettendo una maggiore entrata di Ca^{++} all'interno della cellula; di conseguenza, si ha una più ingente liberazione di mediatore chimico. L'azione della proteinchinasi si manifesta pure mobilizzando il Ca^{++} dai depositi intracellulari, contribuendo anche in questo modo ad aumentare la concentrazione citosolica di questo ione. Ancora sconosciuti nella maggior parte sono gli eventi molecolari responsabili della sensibilizzazione a lungo termine (*linee tratteggiate*). È certo comunque che l'elevato aumento di AMP ciclico in seguito a ripetute applicazioni di serotonina produce un aumento della risposta del motoneurone, che dura oltre le 24 ore. In qualche modo quindi il c-AMP agisce sull'espressione genica. Probabilmente anche in questo caso è necessario l'intervento della proteinchinasi. Quali siano queste proteine sintetizzate e quali funzioni esercitino non è ancora noto. Di certo le modificazioni morfologiche, per esempio l'aumento di zone attive (*triangolini in nero*), viste a carico dei neuroni sensoriali di animali che presentano sensibilizzazione a lungo termine, richiedono la presenza attiva di una sintesi proteica.

In questa coltura (ripresa al microscopio dopo cinque giorni di crescita) di neuroni componenti il riflesso di retrazione della branchia, il corpo cellulare più grosso (circa 200 micrometri) appartiene al motoneurone L7 che innerva parte della muscolatura della branchia. Le tre cellule più piccole sono meccanocettori del raggruppamento LE, che rispondono a stimoli tattili sulla cute del sifone.

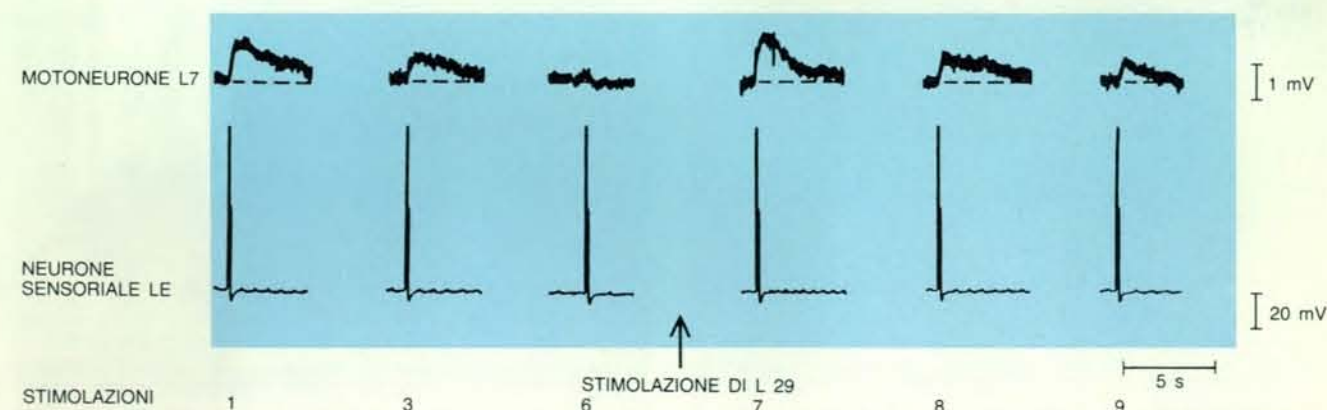
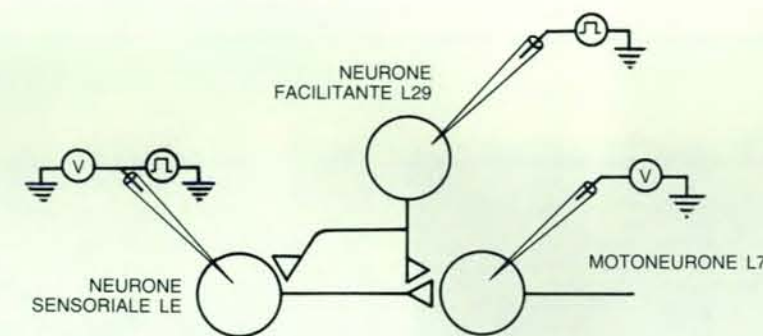
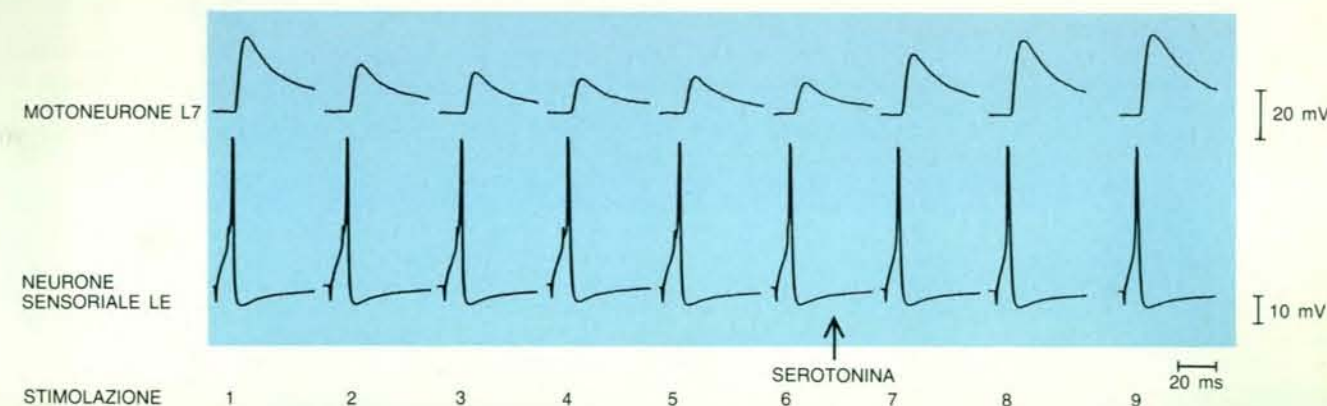
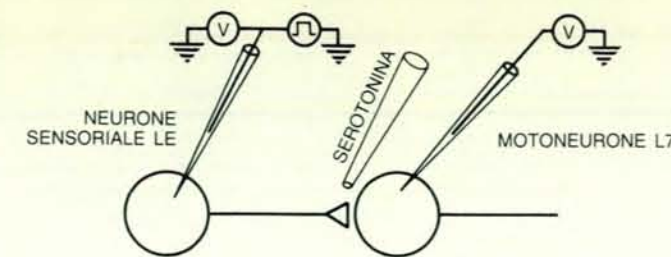
Sigmund Freud. Sfruttando la semplicità del modello sperimentale, cioè il riflesso di retrazione del sifone e della branchia, Eric Kandel e i suoi collaboratori sono riusciti a dare una dimostrazione sperimentale di questa ipotesi affascinante.

Si è detto in precedenza che sia l'assuefazione sia la sensibilizzazione possono durare parecchi giorni. La persistenza della modificazione comportamentale dipende dalla durata del protocollo di stimolazione. Nel caso della sensibilizzazione, una singola scossa elettrica produce un aumento della retrazione della branchia che dura qualche ora, mentre quattro o cinque scosse distanziate di un giorno l'una dall'altra producono un aumento della risposta che dura per settimane. C'è una differenza sostanziale tra i due processi? Oppure sono fasi differenti del medesimo meccanismo? Abbiamo già accennato che le prime osservazioni degli psicologi sperimentali e la casistica neurologica lasciano ritenere che la memoria sia costituita da due componenti distinte: l'una, la memoria a breve termine, facilmente distruttibile; l'altra, la memoria a lungo termine, stabile. Una possibile differenza potrebbe consistere nel fatto che il processo a lungo termine coinvolga un diverso circuito neuronale. Esperimenti di Vincent Castellucci, Tom Carew, William Frost, Craig Bailey e Mary Chen della Columbia University hanno escluso questa possibilità nelle modificazioni comportamentali del riflesso di retrazione della branchia di *Aplysia*.

Nell'analisi dei fattori che differenziano i processi di apprendimento a breve e a lungo termine si è rivelata particolarmente utile la possibilità di ricostruire in coltura parte del circuito neuronale alla base del riflesso di retrazione della branchia di *Aplysia californica*.

Uno di noi (Samuel Schacher) è riuscito a ricostruire in coltura la componente essenziale del riflesso. Con fine tecnica chirurgica si rimuovono dal ganglio addominale alcuni neuroni sensoriali del gruppo LE e il motoneurone L7, il quale è facilmente distinguibile dalle altre cellule nervose per posizione e pigmentazione. Mettendoli in stretta vicinanza in una capsula di Petri contenente un opportuno mezzo di coltura essi formano alcune sinapsi sensomotorie. L'arco riflesso è ridotto alla pura componente monosinaptica. I primi contatti sinaptici rilevabili elettrofisiologicamente si formano nell'arco di 24-48 ore. Questo processo di sinaptogenesi dura tre o quattro giorni fino al raggiungimento della condizione di stabilità che può durare qualche giorno. Successivamente incominciano a manifestarsi alcuni processi degenerativi, che portano alla progressiva scomparsa del contatto sinaptico funzionale e poi alla morte cellulare.

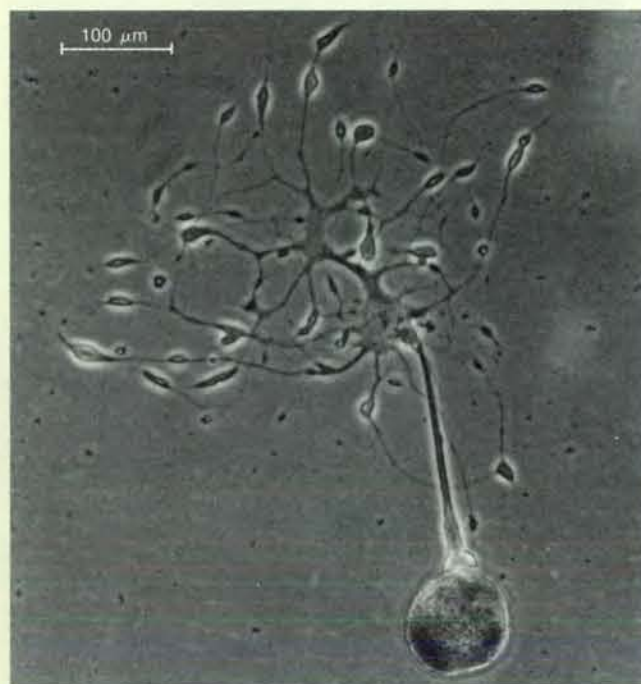
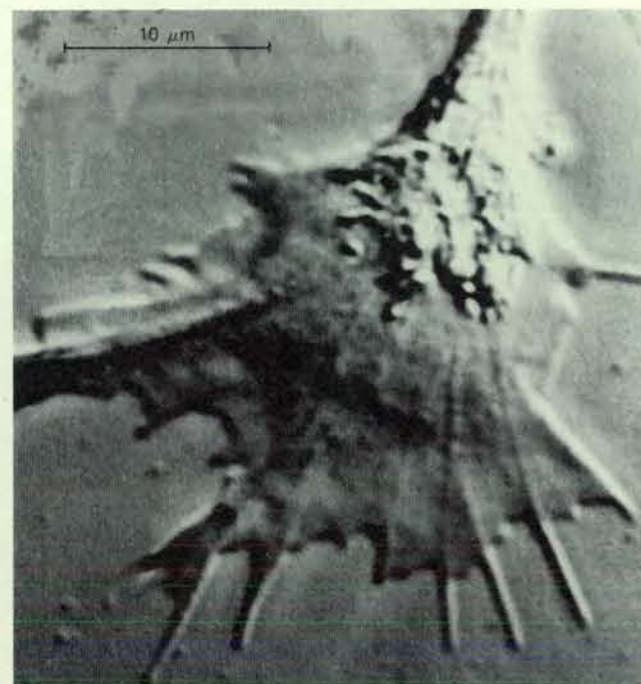
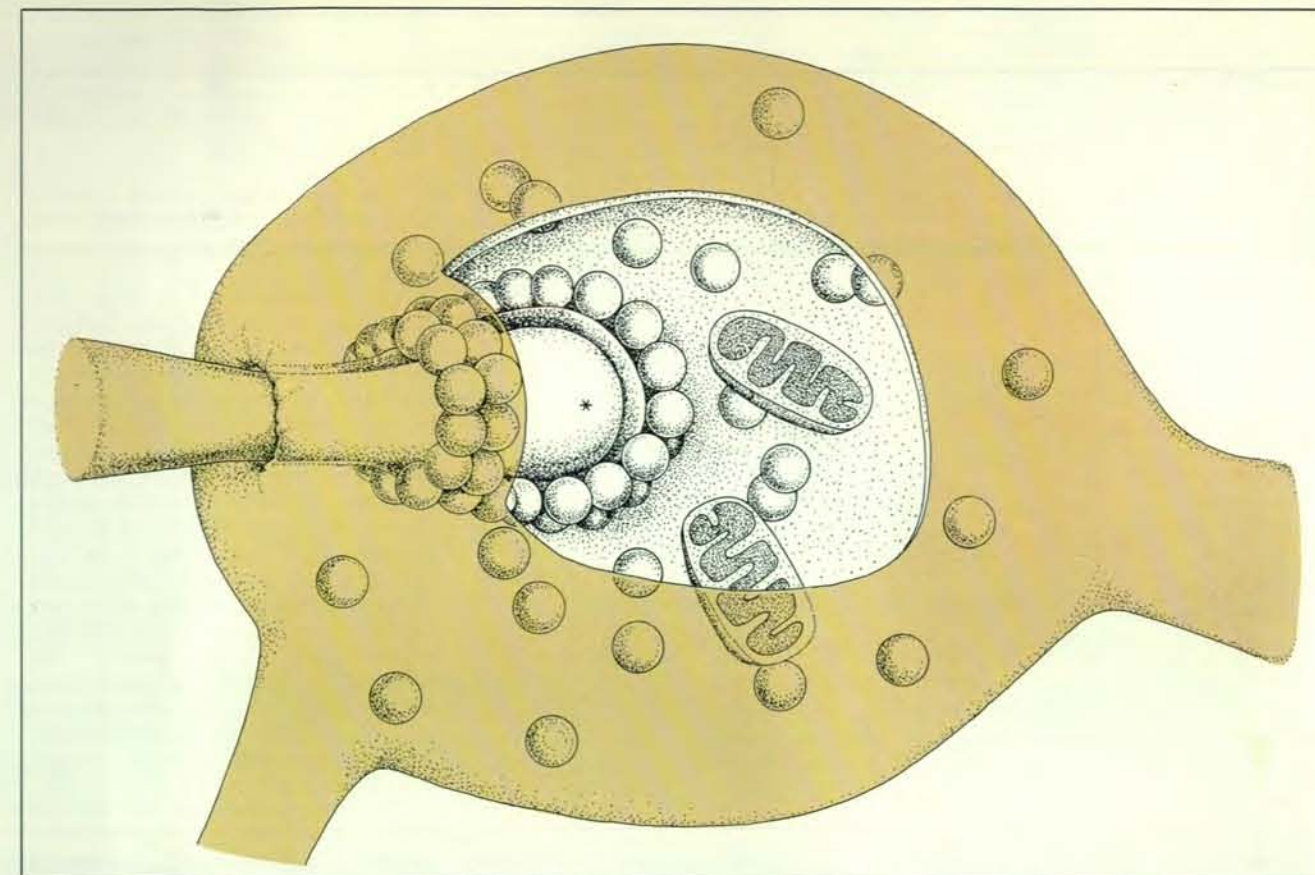
Il semplice circuito in coltura, sopra descritto, presenta le modulazioni della trasmissione sinaptica a breve termine che sono state rilevate nel ganglio addominale? Le stimolazioni tattili della cute del sifone vengono sostituite da stimolazioni elettriche intracellulari del neurone sensoriale; invece dell'ampiezza della contrazione del sifone viene misurato il potenziale postsinaptico eccitatorio (EPSP) del motoneurone. Se si stimola ripetutamente, ogni 20-30 secondi, il neurone sensoriale, si osserva che l'ampiezza dell'EPSP si riduce progressivamente con un decorso temporale del tutto sovrapponibile a quello della contrazione della branchia nell'animale intatto in seguito alla stimolazione ripetuta del sifone. Anche la sinapsi sensomotoria coltivata *in vitro* si «abitua», così come la risposta riflessa *in vivo*; inoltre l'abbassamento del potenziale postsinaptico del motoneurone dura alcuni minuti proprio come la diminuzione della risposta comportamentale. Si parla in questo caso di depressione omosinaptica a breve termine.

Ci siamo chiesti, allora, se il circuito ricostruito in coltura presentasse anche il correlato cellulare della sensibilizzazione, vale a dire la facilitazione sinaptica. La sensibilizzazione a breve termine *in vivo* si ottiene sottoponendo la coda dell'animale a una scossa elettrica. *In vitro* siamo ricorsi, invece, a uno dei presunti mediatori utilizzati dai neuroni facilitanti: la serotonina. L'aggiunta al mezzo di coltura di 2 millimetri cubi di serotonina alla concentrazione di 2 micromoli riporta l'ampiezza dell'EPSP, ridotta dalle precedenti ripetute stimolazioni del neurone sensoriale, ai valori iniziali; questo incremento dura qualche minuto. In collaborazione con Robert

In coltura, il circuito elementare del riflesso di retrazione della branchia presenta le modalità di modulazione sinaptica che costituiscono i correlati cellulari dell'assuefazione e della sensibilizzazione, rispettivamente depressione e facilitazione sinaptica. Nel disegno in alto il circuito è costituito dalla coppia neurone sensoriale e motoneurone. Il neurone sensoriale LE è stimolato ogni due secondi in modo che ogni stimolo produca un solo potenziale d'azione. Il potenziale postsinaptico eccitatorio (EPSP) del motoneurone L7 declina progressivamente (depressione omosinaptica). L'aggiunta di serotonina tra il sesto e il settimo stimolo potenzia l'EPSP ridotto dalle precedenti stimolazioni (facilitazione eterosinaptica). L'intervallo tra l'ottavo e il nono stimolo è stato aumentato a quattro minuti in modo da permettere un accurato lavaggio della serotonina; ciononostante la risposta rimane ancora aumentata. Nel disegno in basso il circuito è costituito da un arco trineuronale comprendente il neurone sensoriale, il motoneurone e uno dei neuroni facilitanti (L29) di cui peraltro non si conosce il mediatore chimico, ma che è stimolato dalla scossa elettrica impartita alla coda. La facilitazione eterosinaptica è analoga a quella ottenuta mediante serotonina. La risposta del motoneurone è stata progressivamente depressa, fino alla scomparsa, tramite ripetute stimolazioni del neurone sensoriale (sono rappresentate le stimolazioni 1, 3 e 6). Nell'intervallo fra la sesta e la settima stimolazione si è stimolato il neurone facilitante così da ottenere una serie di potenziali d'azione ad alta frequenza per simulare l'effetto provocato dalla stimolazione elettrica della coda e indurre la liberazione di neuromodulatore.

Hawkins, abbiamo ottenuto lo stesso risultato con un esperimento più elegante, eseguito su un circuito trineuronale. Il neurone aggiunto alla componente sensomotoria del riflesso è il neurone L29, uno dei neuroni facilitanti attivato dalla stimolazione elettrica della coda. Nel ganglio addominale questa cellula nervosa è in contatto sinaptico con i neuroni sensoriali del gruppo LE. In coltura la stimolazione intracellulare di L29 produce, come la serotonina, una facilitazione della trasmissione a livello della sinapsi sensomotoria.

Francesco Belardetti, ora all'Università del Texas a Dallas, mediante registrazioni effettuate su cellule sensoriali in coltura ha dimostrato che la serotonina, come nel ganglio, esercita la sua azione di modulazione della trasmissione sinaptica chiudendo il canale *S* per il potassio e aumentando, in questo modo, la durata del potenziale d'azione. Egli ha effettuato questi esperimenti non solo utilizzando il corpo cellulare del neurone, ma registrando l'attività elettrica nella parte terminale del neurone in crescita che è detta cono di crescita. Questa struttura appare al microscopio ottico come un'espansione laminare dotata di numerosi prolungamenti filiformi. Le sue discrete dimensioni nel neurone sensoriale *in vitro* hanno permesso di osservare l'azione della serotonina nel distretto neuronale che è considerato la forma immatura della terminazione presinaptica.

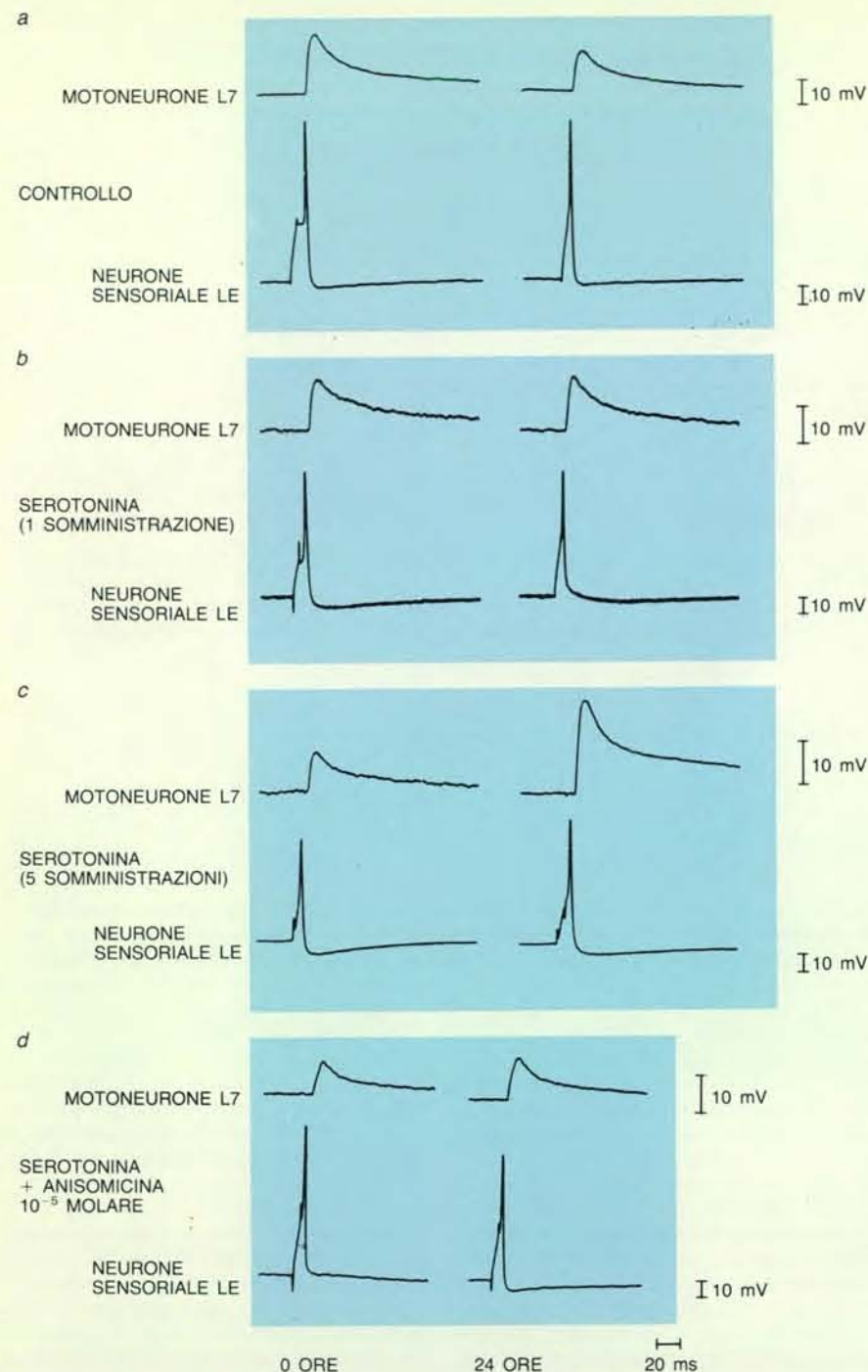
In base a questi risultati abbiamo potuto concludere che, *in vitro*, il circuito semplificato del riflesso di retrazione della branchia si comporta come *in vivo*, dove la modulazione della liberazione del trasmettitore chimico da parte del neurone presinaptico è responsabile della sensibilizzazione a breve termine. Sfruttando i vantaggi sperimentali del circuito *in vitro* - registrare dalle stesse cellule a distanza di 24 ore e avere un miglior controllo, rispetto al ganglio intatto, dell'ambiente biochimico cellulare - abbiamo potuto cominciare a studiare i meccanismi responsabili della sensibilizzazione a lungo termine. Una differenza tra sensibilizzazione a breve termine e sensibilizzazione a lungo termine è di carattere morfologico. Craig Bailey e Mary Chen della Columbia University hanno analizzato al microscopio elettronico la morfologia delle terminazioni dei neuroni sensoriali LE di animali che erano stati sottoposti a un addestramento di sensibilizzazione a lungo termine. Essi hanno iniettato nel neurone sensoriale la perossidasi di rafano, una sostanza comunemente utilizzata in istologia per visualizzare, dopo opportuno trattamento, la cellula iniettata e distinguerne facilmente le strutture fra quelle appartenenti ad altri neuroni. Bailey e Chen hanno così potuto analizzare le terminazioni presinaptiche del neurone sensoriale, le quali, per via del loro aspetto globulare, sono chiamate varicosità. Hanno individuato due tipi di varicosità: quelle dotate di molte vescicole e zone attive e quelle con poche vescicole e mancanti di zone attive. Le prime sono vere e proprie terminazioni presinaptiche. L'altro gruppo può essere considerato una forma non ancora specializzata delle precedenti.

In base alle ricostruzioni di Bailey e Chen in esemplari di *Aplysia* di controllo, circa il 40 per cento delle varicosità era dotato di zone attive. Negli animali sensibilizzati a lungo termine questa percentuale saliva al 70 per cento; inoltre aumentava anche il numero delle vescicole. Per contro, negli animali in cui era

Il neurone sensoriale del raggruppamento LE, qui sopra, è stato fotografato dopo quattro giorni di crescita in coltura. Il diametro del corpo cellulare è di circa 80 micrometri. Quando fu isolato dal ganglio, il neurone era dotato solo della parte iniziale dell'assone senza alcun prolungamento. Nel processo di crescita *in vitro* dapprima si è formato il cono di crescita centrale, da cui sono scaturiti innumerevoli prolungamenti che, a loro volta, hanno formato altri coni di crescita. La fotografia a destra rappresenta l'immagine di un cono di crescita ottenuta mediante videomicroscopia digitale, una tecnica che permette l'osservazione dal vivo e, tramite l'uso del calcolatore, un'alta risoluzione dei dettagli. Questa struttura specializzata del neurite in crescita fu inizialmente descritta dal neurobiologo Santiago Ramón y Cajal. Il cono di crescita è la sede dove viene aggiunto il nuovo materiale necessario all'allungamento del neurite. Esso è estremamente mobile e questa mobilità è stata associata alla funzione di esplorazione dell'ambiente circostante in modo da dirigere l'assone in crescita verso il bersaglio. Le espansioni digitiformi sono i filopodi, mentre le lamine piatte a esse interposte sono i lamellipodi. La funzione di questi ultimi è ancora poco chiara; si ritiene invece che i filopodi, aderendo al substrato, esercitino una trazione sul neurite, facilitando la giustapposizione di nuovo materiale necessario per la crescita. La microfotografia è stata eseguita al Center for Neurobiology and Behavior della Columbia University da Daniel J. Golberg e Donald W. Burmeister.

La ricostruzione tridimensionale della terminazione sinaptica di una cellula sensoriale (varicosità) è stata effettuata da Craig Bailey. Le vescicole sinaptiche, in ognuna delle quali è contenuto un quanto, sono addensate sulla superficie interna specializzata (zona attiva) della membrana presinaptica, la quale forma una tasca che accoglie una spina dendritica del neurone postsinaptico (*asterisco*).

stata indotta una sensibilizzazione a breve termine, non si notava alcuna differenza rispetto al gruppo di controllo. Le osservazioni fatte indicano che, complessivamente, il numero di sinapsi attive aumenta negli animali sensibilizzati a lungo termine. In questi animali i due ricercatori hanno pure descritto un aumento della arborizzazione terminale dell'assone della cellula sensoriale. Vi sono, quindi, evidenti differenze morfologiche tra le due forme di sensibilizzazione. Queste modificazioni della struttura presinaptica possono rappresentare il substrato anatomico del consolidamento della memoria a lungo termine.

Vi è un'altra differenza fra la memoria a breve termine e la memoria a lungo termine: una serie di ricerche iniziate negli anni cinquanta e condotte sui vertebrati superiori ha evidenziato che, mentre la prima è indipendente dalla sintesi proteica, la seconda non si instaura se durante la fase di apprendimento viene bloccata la sintesi sia delle proteine sia dell'RNA. In questi esperimenti, la somministrazione dei farmaci bloccanti avveniva per lo più per via sistemica, per esempio endovenosa; inoltre l'analisi dei processi di memoria era effettuata esclusivamente valutando le modificazioni del comportamento. Queste modalità sperimentali rendono difficile l'interpretazione dei risultati per due motivi: innanzitutto, gli agenti farmacologici utilizzati per bloccare la sintesi proteica possono avere effetti tossici che alterano le buone condizioni dell'animale. In secondo luogo, i farmaci possono influire su altri processi, come per esempio l'attenzione e lo stato di veglia, provocando modificazioni non specifiche del comportamento.

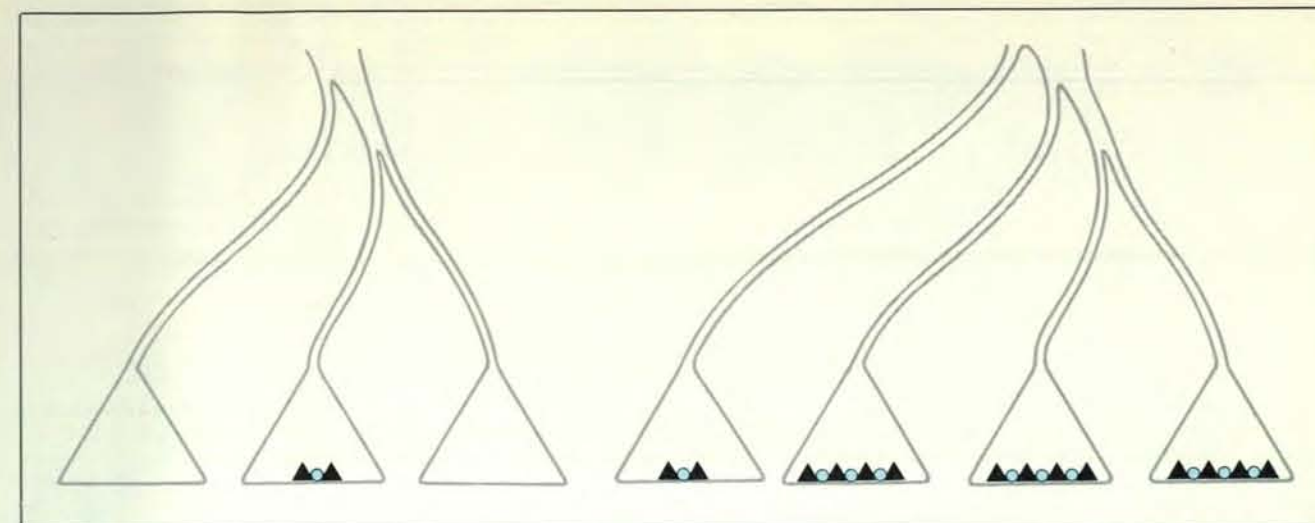
L'impiego del circuito semplificato del riflesso di retrazione della branchia ci ha permesso di superare queste obiezioni. La sinapsi sensomotoria mostrava una facilitazione a breve termine della trasmissione sinaptica; dovevamo prima ottenere un aumento della risposta postsinaptica che durasse almeno un giorno per parlare di plasticità sinaptica a lungo termine. Per produrre una sensibilizzazione a lungo termine della risposta di retrazione della branchia di *Aplysia* sono sufficienti quattro o cinque stimolazioni elettriche della coda, una ogni 30 minuti; l'aumento della risposta dura qualche giorno prima di ritornare ai valori di controllo. Per ottenere facilitazione a lungo termine della trasmissione a livello della sinapsi sensomotoria *in vitro* si è utilizzata di nuovo la serotonina, come succedaneo dello stimolo sensibilizzante. Inizialmente, abbiamo misurato l'ampiezza dell'EPSP del motoneurone; successivamente, in analogia alle cinque stimolazioni sensibilizzanti della coda, abbiamo fatto uso di un protocollo consistente di cinque brevi applicazioni di serotonina, intervallate da risciacqui con liquido di perfusione, della durata di 30 minuti. Dopo questo trattamento, le colture venivano poste nell'incubatore e riesaminate il giorno successivo. Il valore dell'EPSP misurato a distanza di un giorno mostrava un aumento medio del 70 per cento rispetto al valore ottenuto prima del trattamento. Questo dato è interessante per due motivi: 1) indica che nel circuito in coltura è possibile indurre una facilitazione sinaptica di lunga durata; 2) dimostra che è possibile indurre la modificazione sinaptica a lungo termine con il medesimo mediatore chimico che produce facilitazione a breve termine.

Eravamo così in condizione di valutare se il blocco della sintesi proteica impediva l'instaurarsi della facilitazione a lungo termine. Con gli inibitori della sintesi delle proteine e dell'RNA (anisomicina, emetina, actinomicina D e alfa-

I tracciati si riferiscono alla facilitazione eterosinaptica a lungo termine in coltura e al suo blocco da parte dell'anisomicina, uno degli inibitori della sintesi proteica utilizzati. In *a* è rappresentato il controllo. La cellula sensoriale viene stimolata tramite un elettrodo intracellulare così da produrre un solo potenziale d'azione (LE); contemporaneamente viene registrato il potenziale postsinaptico nel motoneurone (L7) (colonna corrispondente alle 0 ore). Immediatamente dopo il circuito in coltura riceve cinque ripetute applicazioni di acqua marina artificiale. A distanza di 24 ore, si nota solo una diminuzione trascurabile della risposta (colonna delle 24 ore). Anche una sola applicazione di serotonina (*b*) non produce alcun effetto rilevabile a distanza di 24 ore, così come una singola stimolazione elettrica della coda causa un aumento della retrazione della branchia che dura solo qualche minuto. Cinque ripetute somministrazioni di serotonina (*c*) al contrario producono un aumento di circa il doppio della risposta 24 ore dopo il trattamento serotoninergico. Se, invece, si blocca la sintesi delle proteine con l'impiego di anisomicina (alla concentrazione di 10 micromoli che bloccano più del 90 per cento della sintesi proteica) per tutta la durata della somministrazione di serotonina, non si osserva più, il giorno seguente, l'incremento della risposta del motoneurone (*d*). Tuttavia l'anisomicina, come gli altri inibitori della sintesi proteica che sono stati utilizzati negli esperimenti, non ha alcun effetto sulla facilitazione eterosinaptica a breve termine e sulla «funzionalità» dei neuroni del circuito *in vitro*.

-amanitina) da noi impiegati durante il ripetuto trattamento con serotonina, abbiamo sempre osservato la mancanza di formazione della facilitazione sinaptica a lungo termine, mentre rimaneva inalterata quella a breve termine.

Sulla base di questi risultati si può affermare che la presenza della sintesi proteica nelle cellule nervose è una prerogativa essenziale per ottenere le modificazioni sinaptiche a lungo termine. Desideriamo sottolineare che l'inibizione della sintesi proteica, per essere efficace, deve avvenire durante il trattamento con serotonina. Se, per esempio, la sintesi di nuove proteine viene bloccata per lo stesso periodo di tempo ma prima (oppure dopo) il trattamento, non si impedisce la formazione della facilitazione a lungo termine. Le proteine indispensabili per la formazione della facilitazione a lungo termine si formano quindi durante il trattamento con serotonina. Questo concetto di finestra temporale entro la quale la sintesi proteica deve essere attiva affinché si possa avere il consolidamento dell'informazione era già emerso dai risultati degli esperimenti condotti sui vertebrati, ai quali si è accennato in precedenza.

Dove avvengono questi eventi molecolari responsabili della facilitazione sinaptica a lungo termine? Vi sono due buone ragioni per ritenere che essi abbiano luogo nel neurone presinaptico: la prima è di ordine morfologico, e ci riferiamo alle modificazioni strutturali delle terminazioni presinaptiche, descritte da Bailey e Chen; l'altra è di ordine elettrofisiologico. Come abbiamo brevemente descritto in precedenza, l'EPSP risulta dalla sommazione dei potenziali postsinaptici in miniatura prodotti dall'azione di un quanto di trasmettitore sulla membrana postsinaptica. Un aumento dell'EPSP, come quello da noi osservato nella facilitazione a lungo termine, potrebbe risultare da un fattore presinaptico (un aumento del numero di quanti liberati per ogni potenziale d'azione) oppure da un fattore postsinaptico (un aumento della sensibilità del recettore postsinaptico per cui ciascun quanto produce un potenziale in miniatura più ampio). Per scegliere tra queste due possibilità si fa uso della cosiddetta analisi quantica della trasmissione sinaptica.

Nicholas Dale che lavora attualmente alla Columbia University, utilizzando questo metodo sperimentale nel circuito in coltura, ha dimostrato che l'aumento dell'EPSP che si osserva nella facilitazione a lungo termine è prodotto da un aumento del numero di quanti liberati dal neurone sensoriale. Il risultato ottenuto da Dale è in accordo con l'osservazione di Bailey che, nell'*Aplysia* sensibilizzata a lungo termine, aumenta il numero di vescicole sinaptiche a livello delle terminazioni del neurone sensoriale. Dall'insieme di questi risultati si deduce che il processo di apprendimento può dare inizio a due tracce mnemoniche nello stesso sito anatomico, mettendo in moto due meccanismi molecolari ben distinti: una modificazione conformazionale (in seguito a fosforilazione) di proteine preesistenti nel caso della memoria a breve termine; la sintesi di nuove proteine in quella a lungo termine.

I cambiamenti morfologici cui vanno incontro i neuroni sensoriali nella sensibilizzazione a lungo termine sono: aumento del numero dei prolungamenti terminali e delle varicosità; aumento del numero di varicosità con zone attive (*triangolini in nero*); incremento di superficie della zona attiva. A sinistra, la condizione di controllo; a destra, quella della sensibilizzazione a lungo termine.

Non si conoscono ancora quali siano i segnali intracellulari che, partendo dal recettore per la serotonina sulla membrana del neurone sensoriale, attivano regioni specifiche del codice genetico; si possono fare alcune considerazioni. Sono state descritte, per esempio, classi di proteine neuronali implicate sia nella trascrizione dell'informazione genica dal DNA all'RNA sia nella traduzione della sequenza nucleotidica dell'RNA nella catena polipeptidica. L'azione di tali proteine è regolata dallo stato di fosforilazione. Il prolungato aumento di AMP ciclico che si verificherebbe durante le ripetute stimolazioni sensibilizzanti potrebbe attivare, tramite la proteinchinasi, queste proteine che regolano l'espressione genica. Non è ancora noto quali siano questi prodotti genici; alcuni di essi potrebbero essere implicati nella formazione dei nuovi contatti sinaptici che caratterizzano la memoria a lungo termine e che forse sono responsabili del suo consolidamento.

La plasticità sinaptica a lungo termine sembra avere una certa analogia con lo sviluppo del sistema nervoso durante il quale si ha crescita neuronale e presenza di sintesi proteica. Il circuito del riflesso di retrazione della branchia è certo predeterminato dai processi di sviluppo, ma può essere modificato dall'apprendimento. Una di queste modificazioni consiste nella formazione di nuovi e più efficaci contatti sinaptici; non pare quindi affatto sorprendente che il processo mnemonico a lungo termine utilizzi gli stessi meccanismi propri dell'ontogenesi del sistema nervoso.

Siamo ben lungi dal riuscire a spiegare i meccanismi della memoria e dell'apprendimento umano, ma fenomeni di rimodellamento sinaptico simili a quelli osservati in *Aplysia* avvengono anche nel sistema nervoso di vertebrati vicini all'uomo nella scala zoologica. William Greenhough dell'Università dell'Illinois ha descritto che, in macachi allevati in colonie e in ambienti ricchi di stimoli, si assiste a una maggior crescita e arborizzazione nel cervelletto delle cellule di Purkinje, una struttura del sistema nervoso centrale che si ritiene implicata nell'apprendimento motorio. In un'altra serie di esperimenti, Greenhough ha osservato che, in animali adulti i quali hanno imparato ad attraversare complicati labirinti per raggiungere il cibo, ha luogo un aumento nel numero dei contatti sinaptici e nella arborizzazione dendritica a carico dei neuroni della corteccia cerebrale. È incoraggiante osservare che, in *Aplysia*, i meccanismi alla base di semplici forme di apprendimento e memoria sono analoghi a quelli descritti da altri ricercatori in diverse specie di invertebrati; inoltre, non sono peculiarità di questi modelli sperimentali. Al contrario, la neuromodulazione della trasmissione sinaptica, le reazioni molecolari a cascata, iniziate dai secondi messaggeri (come l'AMP ciclico), gli effetti biochimici e biofisici della fosforilazione di classi di proteine sono tutti meccanismi generali conservatisi nel processo della filogenesi e diffusi in tutto il regno animale. Per esempio Arey Routtemberg della Northwestern University ha dimostrato, nel macaco, la fosforilazione di proteine che si ritengono associate a modificazioni sinaptiche in aree della corteccia cerebrale, come le aree visive e l'ippocampo, nelle quali si ritiene abbiano luogo processi di memoria.

Allo stato attuale non possiamo né affermare né escludere che la memorizzazione di una poesia necessiti dei meccanismi molecolari descritti. È probabile, però, che la distinzione tra apprendimento e memoria negli invertebrati e nei vertebrati possa perdere di significato via via che aumenta la nostra comprensione di questi processi. L'impiego delle tecniche proprie della biochimica e della biologia molecolare, associate alle tecniche delle scienze psicologiche, nell'analisi dei cambiamenti indotti nel sistema nervoso centrale dall'ambiente e dall'esperienza può colmare questo divario.

BIBLIOGRAFIA

ROBUSTELLI F., *Memoria e apprendimento* in «Le Scienze» n. 65, gennaio 1974.

KANDEL E. R., *Piccoli sistemi di neuroni* in «Le Scienze» n. 135, novembre 1979.

KANDEL E. R. e SCHWARTZ J. H., *Molecular Biology of Learning: Modulation of Transmitter Release* in «Science», 218, n. 4571, 1982.

ALKON D., *L'apprendimento in un mollusco marino* in «Le Scienze» n. 181, settembre 1983.

MONTAROLO P. G., GOELET P., CASTELLUCCI V., MORGAN J., KANDEL E. R. e SCHACHER S., *A Critical Period of Macromolecular Synthesis in Long-term Heterosynaptic Facilitation in* Aplysia in «Science», 234, n. 4781, 1986.

MISHKIN M. e APPENZELLER T., *L'anatomia della memoria* in «Le Scienze» n. 228, agosto 1987.

Metodi di rilevamento della deformazione crostale

Negli Stati Uniti occidentali la crosta si sta deformando per la collisione della zolla del Pacifico con la zolla nordamericana; la misurazione diretta dell'entità del fenomeno è oggi possibile con tecniche di geodesia spaziale

di Thomas H. Jordan e J. Bernard Minster

In molte zone degli Stati Uniti occidentali la crosta terrestre si sta attivamente deformando. Lungo la faglia californiana di San Andreas, che è sicuramente la più famosa linea su cui avvengono terremoti di grave entità nell'America Settentrionale, la deformazione prende la forma di uno scorrimento orizzontale: la crosta a ovest della faglia sta spostandosi verso nord-ovest rispetto alla crosta che si trova sull'altro lato. A est della faglia di San Andreas vi è un'altra area geologicamente attiva, la Basin and Range Province, una regione ampia 1000 chilometri che include il Great Basin del Nevada e dello Utah occidentale. Qui la crosta è in via di lacerazione, probabilmente a causa di forze che hanno origine nel mantello terrestre (lo strato immediatamente al di sotto della crosta). A ovest della faglia di San Andreas vi è una zona geologicamente attiva molto più stretta, che comprende la catene montuose costiere della California e il margine continentale. Qui la deformazione si manifesta nelle montagne, che tuttora sono in corso di sollevamento a opera delle complesse forze compressive, e nei movimenti tellurici.

Queste deformazioni, per quanto possano sembrare del tutto distinte le une dalle altre, possono essere viste come elementi di un movimento a scala molto più grande: il moto relativo di due immense porzioni della superficie terrestre, vale a dire la zolla del Pacifico e la zolla nordamericana. La zolla del Pacifico sta scorrendo verso nord-ovest rispetto a quella nordamericana a una velocità di circa 50 millimetri all'anno. La parte occidentale degli Stati Uniti è presa nel mezzo; è qui che il movimento delle zolle deve risolversi in attività tettonica, cioè con la deformazione della crosta continentale. Quindi le varie deformazioni che hanno luogo negli Stati Uniti occidentali - l'espansione della Basin and Range Province, lo scorrimento orizzontale lungo la faglia di San Andreas e la complessa tettonica della costa californiana - devono, prese nell'insieme, giustificare il movimento relativo totale della zolla del Pacifico e di quella nordamericana.

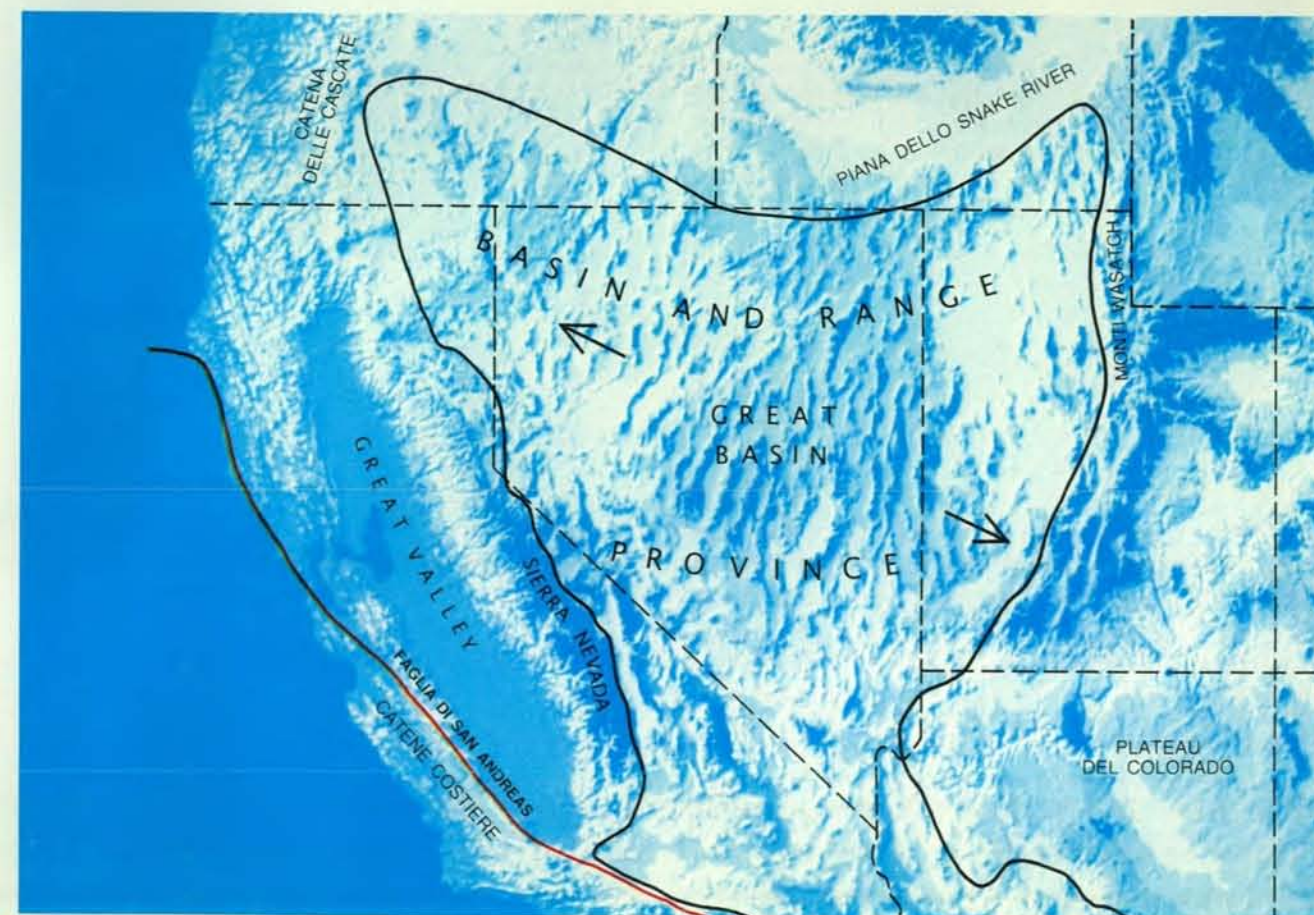
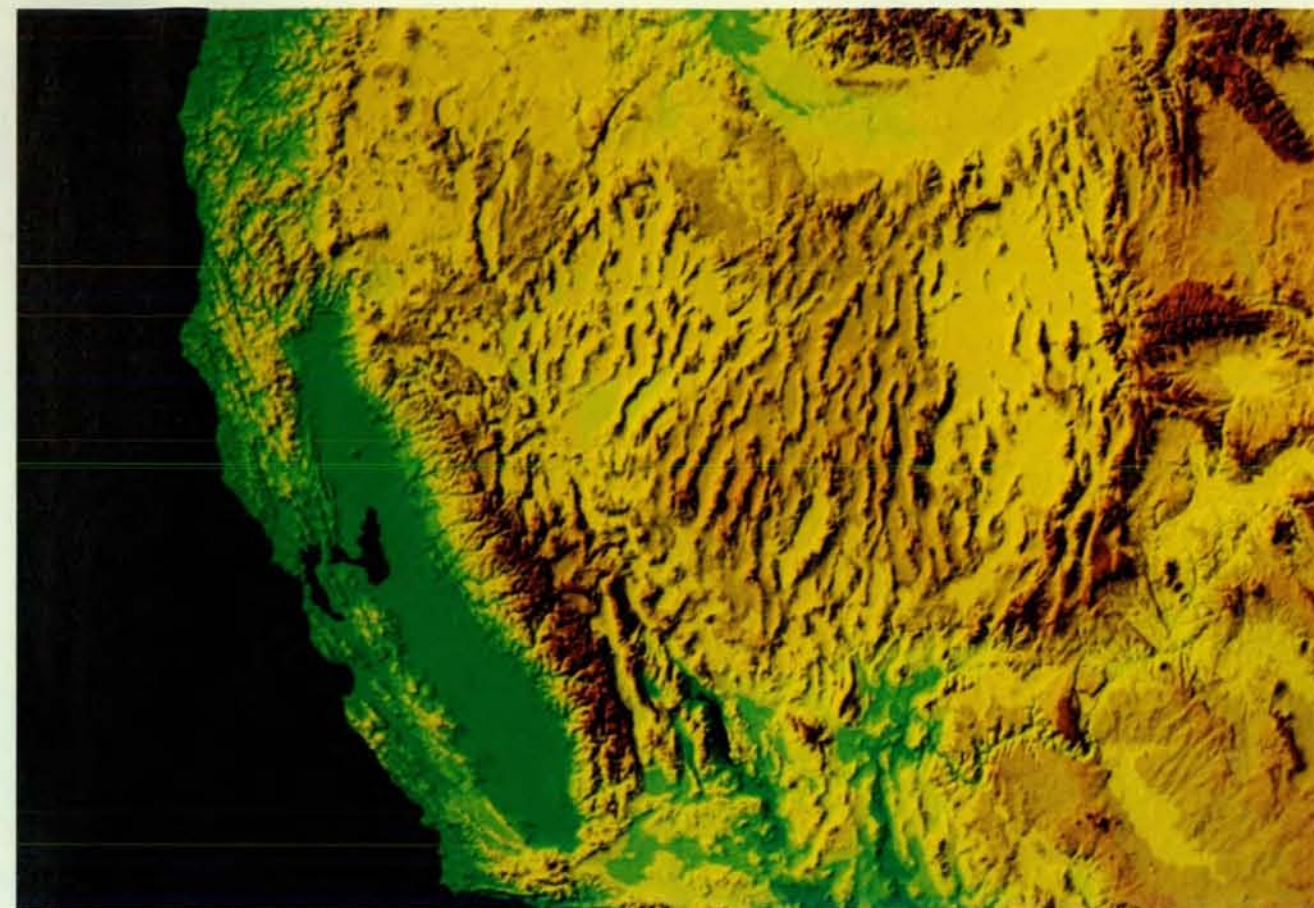
Come è distribuito questo movimento a grande scala nella parte occidentale degli Stati Uniti? Quanta parte di esso si manifesta in ciascuna delle tre principali zone di deformazione?

Il quesito presenta interesse da un punto di vista puramente scientifico, ma nondimeno ha una grande importanza pratica. Al fine di valutare i rischi di terremoti in una regione, i geologi devono per prima cosa comprendere in che modo la regione si stia deformando. Sfortunatamente è talvolta difficile o persino impossibile osservare direttamente la deformazione. Questo è il caso in particolare della regione costiera californiana, dove le manifestazioni geologiche che potrebbero permettere l'interpretazione sono sommerse dalle acque dell'Oceano Pacifico; il rischio sismico in questa regione è al centro di un intenso e annoso dibattito.

Possiamo forse migliorare la nostra comprensione della deformazione che qui ha luogo considerandola come un elemento del più ampio contesto dato dal movimento relativo della zolla del Pacifico e di quella nordamericana.

Per riuscire a comporre questo quadro dobbiamo giungere a una migliore acquisizione dei tassi di deformazione. A questo scopo ci rivolgiamo alla geodesia, la scienza che misura le posizioni sulla superficie terrestre. Il rilevamento geodetico si basa su misurazioni estremamente precise di angoli e distanze per determinare le posizioni relative di punti in una rete; i tassi di deformazione possono essere calcolati determinando di quanto le distanze fra i punti cambiano nel corso del tempo. Per esempio, un moderno geodimetro a laser misura il tempo di percorso di un fascio laser rimandato da un riflettore ad angolo, posto in una stazione lontana, con una precisione superiore al decimiliardesimo di secondo, al fine di ottenere la distanza lungo la linea di vista con una precisione entro il centimetro.

La curvatura della Terra e l'atmosfera limitano queste misurazioni lungo la linea di vista a distanze di poche decine di chilometri, il che è sufficiente per il rilevamento di strette zone di faglia, ma non per studiare i tassi di deformazione su regioni ampie come la parte occidentale degli Stati Uniti. Negli ultimi due decenni sono state sviluppate nuove tecniche che prevedono l'uso di punti di riferi-

La parte occidentale degli Stati Uniti va deformandosi attivamente in diverse regioni. Il Great Basin (*al centro*) si sta espandendo e il processo di espansione dà origine a dorsali (*horst*) e valli tettoniche (*graben*) che decorrono circa perpendicolarmente alla direzione di espansione. Lungo la faglia di San Andreas (*a sinistra*) la deformazione prende la forma di uno scorrimento orizzontale: la crosta a ovest della faglia va spostandosi a nord-ovest rispetto alla crosta che si trova a est. La regione costiera californiana (*in basso a sinistra*) sta pure deformandosi; è in corso una modificazione morfologica causata da terremoti e da piegamenti della crosta, che stanno dando origine a nuove catene montuose. L'immagine in alto è stata generata al calcolatore a partire da una base di dati di più di otto milioni di punti topografici, raccolta dalla National Oceanic and Atmospheric Administration.

mento extraterrestri - satelliti orbitanti intorno alla Terra e radiosorgenti extragalattiche come i quasar - per tenere sotto controllo le deformazioni. Queste tecniche «spazio-geodetiche» permettono di misurare distanze fra punti lontani 1000 chilometri con precisione anche superiore al centimetro. La tecnologia è nuova, ma ha rivoluzionato la scienza della misurazione delle distanze e contribuisce già alla osservazione diretta dei movimenti geologici nella parte occidentale del continente americano.

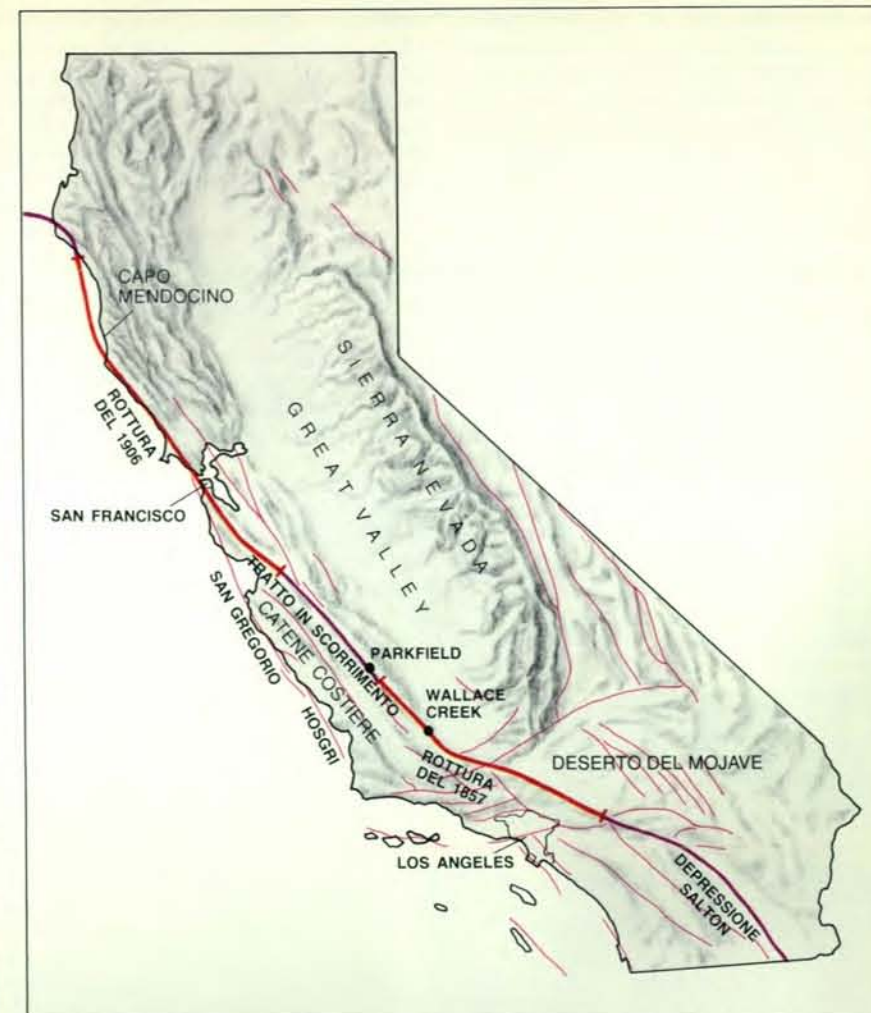
L'idea che esista una stretta relazione tra la deformazione nell'ovest degli Stati Uniti e il movimento a grande scala delle zolle tettoniche ha le sue origini negli studi sulla faglia di San Andreas. Nel famoso terremoto di San Francisco del 1906, una sezione lunga 420 chilometri della faglia subì un dislocamento, causando estesi danni e provocando un disastroso incendio nella città. Recinzioni, strade e ferrovie attraversate dalla faglia furono tagliate e sfalsate, cosicché i segmenti su un lato della faglia non combaciavano più con quelli sull'altro lato. Attente osservazioni del fenomeno condussero Harry Fielding Reid della Johns Hopkins University a formulare quella che ora è nota come teoria del riassestamento elastico (si veda l'articolo *Il movimento del suolo nei terremoti* di David M. Boore in «Le Scienze» n. 114, febbraio 1978). In base a questa teoria la faglia separa blocchi crostali che si muovono a velocità costante l'uno rispetto all'altro; l'attrito sul piano di faglia impedisce che i blocchi si spostino fino a che la tensione accumulata eccede la resistenza delle rocce, ed è allora che avviene una dislocazione quasi istantanea, vale a dire un terremoto.

Il movimento relativo di enormi porzioni di superficie terrestre, la zolla del Pacifico (*a sinistra*) e quella nordamericana (*a destra*), è il meccanismo che innesca terremoti e deformazioni crostali nell'ovest degli Stati Uniti. Lungo la costa occidentale la zolla del Pacifico si muove verso nord-ovest, al tasso di circa 50 millimetri all'anno, rispetto alla zolla nordamericana. Circa due terzi del movimento si risolvono in uno scorrimento orizzontale lungo la faglia di San Andreas. Il movimento rimanente si esprime nell'espansione della Basin and Range Province, che comprende Utah, Nevada e Arizona, nello scorrimento orizzontale lungo altre faglie in California oltre a quella di San Andreas e in altri tipi di deformazione (come quella che dà luogo alla formazione di montagne) sulla costa californiana.

In una pubblicazione fondamentale del 1965, J. Tuzo Wilson, allora all'Università di Toronto, avanzava l'ipotesi che lo scorrimento lungo la faglia di San Andreas fosse provocato dal moto relativo di due delle grandi zolle in cui, come i geologi stavano rendendosi conto in quel momento, è divisa la superficie terrestre. Alla fine degli anni sessanta i geofisici, che stavano tentando di sviluppare il primo modello coerente di tettonica globale, fecero propria l'ipotesi di Wilson che la faglia di San Andreas costituisse il margine fra la zolla del Pacifico e quella nordamericana.

Allora non erano disponibili valutazioni dirette del tasso attuale di scorrimento lungo la faglia, ma il tasso di produzione di nuova crosta nell'Oceano Pacifico (tasso direttamente legato alla velocità delle zolle) suggeriva che le zolle si muovessero l'una rispetto all'altra di circa 60 millimetri all'anno. I geologi si resero conto ben presto, comunque, che se il tasso di scorrimento lungo la faglia di San Andreas fosse stato costantemente così elevato, alcune colate di lava, continue ai due lati della faglia al tempo della loro formazione, circa 10 milioni di anni fa, dovrebbero ora essere sfalsate di 600 chilometri; la dislocazione reale è invece di 250 chilometri.

D'altro canto, il movimento lungo la faglia non si è svolto a velocità costante per 10 milioni di anni e in questo periodo si sono verificati drastici cambiamenti nella geometria delle zolle. Si sono fatti tentativi per far quadrare il lento andamento dei fenomeni geologici con le elevate velocità di spostamento delle zolle, rifacendosi appunto a tali presunti cambiamenti. Per esempio, l'attuale tasso di scorrimento della faglia di San Andreas è forse maggiore del tasso dei fenomeni geologici. Nel 1973 comunque, James C. Savage e Robert O. Burford dello US Geological Survey dimostrarono che esiste di fatto una discrepanza fra il tasso di scorrimento orizzontale della faglia e il movimento relativo delle zolle: precisi rilevamenti effettuati con geodimetri nella California centrale hanno rivelato che tra il 1960 e il 1970 il tasso di scorrimento lungo la faglia di San Andreas è stato di soli 32 millimetri all'anno, con un'incertezza di soli cinque millimetri all'anno.

Circa nello stesso periodo, le incertezze nella descrizione dei movimenti delle zolle andavano via via riducendosi, grazie a modelli che facevano uso sistematico delle grandi masse di dati raccolte dagli oceanografi per tutto il decennio precedente. In collaborazione con Peter H. Molnar, che allora si trovava alla Scripps Institution of Oceanography, realizzammo un modello delle velocità delle zolle mediate sugli ultimi due-tre milioni di anni e lo chiamammo RM1

Le faglie attive, lungo le quali è più probabile che avvengano terremoti, sono numerose in California. La più famosa è la faglia di San Andreas (*in arancione e viola*) lungo la quale hanno avuto luogo molti terremoti di grande intensità, come il famoso terremoto di San Francisco del 1906. Altre faglie attive sono indicate con il tratto sottile in rosso.

(movimento relativo 1). Il modello indicava che la velocità relativa della zolla del Pacifico e di quella nordamericana era all'incirca di 57 millimetri all'anno, velocità che si accordava con le precedenti stime ed era quasi il doppio del tasso di scorrimento orizzontale misurato geodeticamente lungo la faglia di San Andreas. Molnar e Tanya M. Atwater, anch'ella alla Scripps, notarono che, in assenza di dati geologici di avvenute grandi variazioni del tasso di scorrimento negli ultimi milioni di anni, il modo più ovvio per spiegare la differenza fra i valori geodetici e quelli forniti dalla tettonica a zolle era di attribuire il «movimento mancante» a dislocazioni su altre faglie oltre che su quella di San Andreas, in particolare le faglie attive nella Basin and Range Province.

Nei 15 anni che seguirono alla prima scoperta della «discrepanza di San Andreas», le osservazioni che l'avevano messa in evidenza aumentarono di numero e attendibilità. Le stime del tasso di scorrimento orizzontale lungo la faglia di San Andreas, inizialmente fatte da Savage e Burford, hanno retto bene. Inoltre lo studio geologico - il rilevamento in dettaglio e la datazione di suoli alluvionali depositati da un torrente che attraversa la faglia, a opera di Kerry E. Sieh del California Institute of Technology e del compianto Richard H. Jahns della Stanford University - ha dimostrato che il tasso medio di scorrimento durante gli ultimi 13 000 anni è stato di 36 millimetri all'anno, cioè circa uguale a quello attuale. Lo stretto accordo con i dati geodetici è significativo, poiché i dati geologici riflettono gli effetti cumulativi di molti intensi movimenti tellurici e sono pertanto abbastanza indenni da fluttuazioni di breve termine.

Anche le stime delle velocità di spostamento delle zolle hanno avuto un miglioramento. Nel 1978 compilammo una serie «migliorata» di dati oceanografici, al fine di produrre un modello aggiorna-

to del movimento globale delle zolle, chiamato RM2. Sebbene esso incorporasse sostanziali revisioni delle velocità delle zolle calcolate per molte parti del mondo mediante RM1, il valore ricavato per il tasso di movimento relativo fra la zolla del Pacifico e quella nordamericana - 56 millimetri all'anno con un'incertezza di tre millimetri all'anno - era sostanzialmente uguale nei due modelli. Una velocità all'incirca uguale fu stimata in un altro modello globale reso noto nello stesso anno da Clement G. Chase dell'Università del Minnesota.

La discrepanza tra il tasso di scorrimento osservato della faglia di San Andreas e il tasso calcolato di movimento relativo delle zolle rimaneva quindi sui 22 millimetri all'anno, cioè un enorme 40 per cento del movimento delle zolle! Come si poteva spiegare questa discrepanza? Dal 1978 i ricercatori hanno ristretto le loro ipotesi a tre possibilità: una, che il movimento delle zolle stimato nel modello sia troppo alto; la seconda, che il movimento mancante vada accreditato all'espansione della Basin and Range Province; la terza, che il movimento mancante si traduca invece in deformazioni della regione costiera californiana. Queste ipotesi non si escludono a vicenda, anzi, una combinazione di tutte e tre potrebbe meglio adattarsi ai dati di fatto.

La stima ottenuta con il modello RM2 era intrinsecamente incerta a causa delle difficoltà implicite nella stima del movimento della zolla del Pacifico e di quella nordamericana. La valutazione della velocità relativa delle zolle dipende criticamente dalla interpretazione della geologia del fondo marino di un esiguo segmento di dorsale all'imboccatura del Golfo di California, appena a nord della faglia trasforme di Tamayo.

Al tempo dell'elaborazione del modello, su questa zona erano disponibili ben pochi dati. Da allora l'area è stata sottoposta molte volte a rilevamenti e, a partire dai nuovi dati, Charles DeMets e colleghi alla Northwestern University hanno trovato nel 1987 che le zolle si stanno muovendo l'una rispetto all'altra al tasso di circa 48 millimetri all'anno. Questo nuovo valore riduce la discrepanza di San Andreas a circa 15 millimetri all'anno, un valore più piccolo, ma ancora considerevole.

Dove va a finire questo deficit di movimento? Il disavanzo fra il tasso di scorrimento lungo la faglia di San Andreas e il movimento delle zolle è di entità sufficiente a turbare i sonni di un sismologo o di un responsabile della protezione civile. È paragonabile, per esempio, al tasso di scorrimento totale lungo la faglia dell'Anatolia settentrionale o la faglia di Motagua in Guatemala, entrambe segnate da lunghe e funeste storie di attività sismica. Il movimento mancante potrebbe essere distribuito in molte faglie su tutta la parte occidentale degli Stati Uniti, oppure potrebbe essere concentrato in poche faglie o essere assorbito da deformazioni della crosta costanti e asismiche. Queste possibilità hanno evidentemente implicazioni molto diverse riguardo al rischio sismico. Un modo di iniziare ad affrontare il problema è quello di domandarsi quanta parte del movimento mancante abbia luogo nelle regioni a est della faglia di San Andreas, come il Great Basin, e quanta a ovest, lungo la costa californiana.

I nostri strumenti per analizzare il movimento mancante sono i vettori. I vettori, disegnati come frecce, rappresentano il moto; la lunghezza di un vettore corrisponde alla velocità del moto e la direzione verso cui il vettore punta rappresenta la direzione del moto. L'effetto netto di una qualunque combinazione di moti può essere determinato con una somma vettoriale: i vettori vengono disposti in modo che la «coda» di uno coincida con la «testa» dell'altro, dopodiché un terzo vettore viene tracciato a congiungere la coda del primo con la testa del secondo. Questo vettore, che è la somma vettoriale dell'insieme, rappresenta il moto netto.

Nel nostro caso sappiamo che il moto netto - cioè la somma dei contributi dovuti allo scorrimento lungo la faglia di San Andreas, all'espansione che si ha a est della faglia stessa (cioè nella Basin and Range Province) e alle deformazioni che si verificano a ovest, nella regione costiera californiana - dovrebbe coincidere con il movimento relativo della zolla del Pacifico e di quella nordamericana. Sappiamo pertanto che i vettori che rappresentano questi tre tipi di moto, addizionati vettorialmente, dovrebbero dare un vettore uguale, in lunghezza e orientazione, a quello che rappresenta il movimento relativo delle zolle.

Come punto di riferimento per i nostri calcoli vettoriali abbiamo scelto un sito sulla faglia di San Andreas, a 36 gradi di latitudine nord, pochi chilometri a nord del piccolo centro di Parkfield. (Piccolo davvero, in quanto la popolazione non supera le 50 anime, ma famoso nell'ambiente dei sismologi poiché è stato previsto che entro i prossimi cinque anni vi accadrà un terremoto.) In questa parte della California la faglia piega di 41 gradi verso ovest rispetto alla direzione nord. Un paziente osservatore situato sul margine est della faglia in prossimità del punto di riferimento vedrebbe un oggetto di riferimento piazzato sul lato ovest spostarsi verso destra (vale a dire, verso nord-ovest) alla velocità media di circa 34 millimetri all'anno.

La «discrepanza di San Andreas» è la non corrispondenza tra il tasso e la direzione dello scorrimento orizzontale lungo la faglia di San Andreas e il movimento relativo della zolla del Pacifico e di quella nordamericana. Il vettore in viola indica come si muove la zolla del Pacifico rispetto a quella nordamericana. Il vettore arancione mostra la velocità e la direzione di scorrimento lungo la faglia di San Andreas. Il vettore a tratteggio è il movimento «eccedente» che deve essere «assorbito» in altre parti degli Stati Uniti occidentali.

Si può stimare l'espansione del Great Basin a partire dalle variazioni di lunghezza delle linee di base misurate per mezzo dell'interferometria a lunghissima base (VLBI). Con questa tecnica si registra mediante radiotelescopi il «rumore» ad alta frequenza emesso da sorgenti extragalattiche (come i quasar). Si può calcolare la distanza fra i telescopi con la precisione di circa un centimetro confrontando i tempi di arrivo delle onde radio. Le variazioni di lunghezza delle linee di base interferometriche misurate tra il 1980 e il 1985 nell'ambito del Crustal Dynamics Project della NASA sono date in millimetri all'anno. Il tasso di variazione lungo la linea di base tra Onsala e Owens Valley è stato corretto per tener conto del movimento relativo delle zolle nordamericana ed eurasiatica.

Questa informazione - direzione e velocità di scorrimento - ci mette in grado di tracciare il vettore che rappresenta il moto dovuto allo scorrimento lungo la faglia. Il passo successivo consiste nel trovare la direzione e la velocità del movimento che si verifica a est della faglia di San Andreas, movimento che è associato in primo luogo all'espansione della Basin and Range Province. Quindi è abbastanza semplice trovare il vettore rimanente, quello che rappresenta il movimento dovuto alla deformazione lungo la fascia costiera californiana. Si sommano semplicemente i vettori che rappresentano lo scorrimento orizzontale lungo la faglia di San Andreas e l'espansione della Basin and Range Province e si confronta questa somma con il vettore che rappresenta il moto relativo della zolla del Pacifico e di quella nordamericana, derivato da modelli come RM2. Qualsiasi differenza va probabilmente attribuita alla deformazione in atto nella fascia costiera della California. Avremo quindi una descrizione quantitativa di come il movimento delle zolle sia distribuito fra le tre principali zone di deformazione degli Stati Uniti occidentali.

Lo studio geologico della Basin and Range Province ebbe inizio con il lavoro di John Wesley Powell, Grove Karl Gilbert e altri pionieri dello US Geological Survey, che riconobbero l'andamento di base della deformazione circa un secolo prima della scoperta della tettonica delle zolle. I dettagli sono complessi, ma l'elemento predominante della deformazione è l'espansione lungo una direttrice orientata grosso modo da nord-ovest a sud-est. Via via che la crosta si espande, alcuni blocchi crostali cedono e sprofondano, dando luogo a *graben*, fosse tettoniche con andamento circa perpendicolare alla direttrice di espansione. Le dorsali che separano i graben sono alti tettonici, chiamati *horst*. Alle latitudini di Denver e San Francisco, la Basin and Range Province comprende circa 20 strutture a horst e graben, che vanno dal fronte di Wasatch (un importante sistema di faglie che corre in direzione circa nord-sud attraversando Salt Lake City) alle ripide pendici della Sierra Nevada. Gran parte di questa estensione è occupata dal Great Basin, una depressione chiusa, ampia 500 chilometri, originata dallo stiramento e dall'assottigliamento della crosta.

In questa regione vi sono molti modi per trovare la direzione netta di espansione. Le strutture a horst e graben sono pressoché perpendicolari alla direzione in cui i blocchi crostali si allontanano l'uno dall'altro; dalle tracce di scorrimento lungo i piani di faglia si hanno buone indicazioni sulla direzione in cui è avvenuta l'espansione. Ulteriori dati vengono dall'analisi di talune formazioni vulcaniche. Per esempio, i coni di cenere tendono a formarsi lungo linee che seguono fratturazioni verticali perpendicolari alla direzione di espansione. La direzione in cui sta avvenendo lo stiramento della crosta può essere dedotta anche trivellando in profondità e iniettando fluidi ad alta pressione: la crosta tende a fratturarsi lungo piani perpendicolari all'asse di massima tensione.

Nel 1980 Mark L. e Mary Lou Zoback dello US Geological Survey combinarono diverse di queste tecniche per giungere a una descrizione completa delle orientazioni di tensione e di deformazio-

La rappresentazione vettoriale dell'espansione della Basin and Range Province (*in blu*) mostra che il movimento è troppo lento, e per di più nella direzione sbagliata, per giustificare completamente la discrepanza di San Andreas (*in arancione*). Il tasso di espansione è stato ottenuto in gran parte dai dati interferometrici e la direzione soprattutto dai rilevamenti geologici. Le curve chiuse definiscono il campo di variabilità della lunghezza e della direzione dei vettori, entro i livelli indicati di attendibilità delle osservazioni.

ne entro il Great Basin. Scoprirono che la crosta a ovest, tra il Great Basin e la faglia di San Andreas, sta spostandosi rispetto alla parte relativamente stabile del continente nordamericano a est del Great Basin, in una direzione circa 60 gradi ovest in riferimento al nord. Questa è pertanto la direzione del vettore che rappresenta il movimento dovuto all'espansione del Great Basin.

Non sorprende che la valutazione del tasso di espansione effettuata sulla base dei dati geologici si sia dimostrata più difficile che non la stima della direzione di espansione. In una rassegna pubblicata parecchi anni fa abbiamo trovato dati geologici e geofisici tali da corrispondere a tassi di espansione oscillanti tra appena un millimetro all'anno e più di 20 millimetri all'anno.

Le migliori stime geologiche del tasso di espansione vengono dalle nuove tecniche di mappatura della paleosismologia, un giovane ramo delle scienze della Terra che studia le dislocazioni causate da singoli terremoti avvenuti in tempi preistorici. Finora solo poche faglie nella Basin and Range Province sono state studiate in dettaglio, ma sulla base della mappatura preliminare eseguita da Robert E. Wallace dello US Geological Survey siamo giunti alla conclusione che il tasso di espansione nel Great Basin non dovrebbe aver superato i 12 millimetri in media all'anno durante gli ultimi 12 000 anni. Sfortunatamente alcuni degli assunti che rientrano nella stima del tasso di espansione a partire da osservazioni geologiche e geofisiche sono discutibili; le conferme incrociate fra i dati sono poche e il credito che concediamo a queste stime non è molto alto.

La geodesia, da parte sua, dovrebbe essere in grado di fornire una misura diretta, indipendente e precisa del tasso di espansione. Sebbene le misurazioni geodetiche lungo la linea di vista siano limitate a distanze non superiori a qualche decina di chilometri, l'espansione lungo la direttrice da nord-ovest a sud-est nella Basin and Range Province potrebbe, in linea di principio, essere determinata tramite misurazioni ripetute su una rete che copra tutta la regione. Si tratterebbe di un'operazione noiosa e dal costo elevato, essendo necessario il ricorso a molti punti di riferimento intermedi. Inoltre, dato che a ogni passo si accumulerebbero errori, il risultato finale non sarebbe abbastanza attendibile per un modello tettonico.

L'alternativa è la geodesia spaziale. Una tecnica particolarmente vantaggiosa, nota come interferometria a lunghissima base (VLBI), è concettualmente molto semplice (si veda l'articolo *Lo studio della Terra con la VLBI* di William E. Carter e Douglas S. Robertson in «Le Scienze» n. 221, gennaio 1987). Il «rumore» ad alta frequenza emesso da un quasar viene registrato da radiotelescopi in stazioni distanti fra loro e correlato per determinare la differenza tra i tempi di arrivo delle onde radio alle diverse stazioni. Queste differenze di tempi misurano direttamente di quanto una stazione sia più vicina alla sorgente rispetto all'altra. Effettuando una serie di misurazioni su un gruppo di quasar in varie parti del cielo, si può trovare la posizione di ogni stazione nel sistema di riferimento dato dai quasar e determinare come tale posizione vari con il tempo.

Sebbene l'interferometria a lunghissima linea di base abbia mosso i primi passi negli anni sessanta, la tecnica è stata portata alla precisione dell'ordine di un centimetro necessaria per gli studi sulla deformazione crostale non prima del 1980. A quell'epoca una rete globale di stazioni interferometriche era stata ultimata dalla NASA nell'ambito del Crustal Dynamics Project e le osservazioni dei movimenti relativi delle zolle stavano già rendendosi disponibili. Nel giro di cinque anni, i gruppi di lavoro condotti da Thomas A. Clark del Goddard Space Flight Center della NASA e da Thomas A. Herring e Irwin I. Shapiro dello Harvard-Smithsonian Center for Astrophysics furono in grado di raccogliere ed elaborare un numero sufficiente di dati interferometrici per poter valutare i tassi di movimento su linee di base estese fino a comprendere l'Atlantico e il Pacifico; i valori differivano generalmente dai calcoli basati su modelli tettonici del tipo di RM2 di soli pochi millimetri all'anno.

Spronati da questo successo, abbiamo fatto uso dei dati interferometrici per stimare sia la direzione sia il tasso di espansione nella Basin and Range Province. Il tasso di espansione stimato è di nove millimetri all'anno, con un'incertezza di quattro millimetri, e la direzione è di 48 gradi ovest rispetto al nord, più o meno 17 gradi. Questi valori geodetici sono indipendenti dai dati geologici e in buon accordo con essi. I dati interferometrici forniscono un tasso più preciso, mentre i dati geologici danno una migliore valutazione della direzione; nel calcolare il vettore che rappresenta l'espansione nel Great Basin, ci siamo pertanto riferiti principalmente ai dati interferometrici per quanto riguarda la lunghezza del vettore e a quelli geologici per quanto riguarda la direzione.

Il vettore che si ottiene combinando i diversi insiemi di dati indica un tasso di espansione di 10 millimetri all'anno con un'incertezza di due millimetri all'anno e una direzione 56 gradi ovest rispetto al nord con un'incertezza di più o meno 10 gradi. È un movimento troppo lento e nella direzione sbagliata per giustificare del tutto la discrepanza di San Andreas. Effettuando una semplice somma vettoriale, troviamo che la deformazione nella regione costiera californiana si esprime in un movimento netto di circa nove millimetri all'anno, in una direzione circa 14 gradi est rispetto al nord.

Tutto ciò, dunque, ci dice come è distribuito il movimento delle zolle nella parte occidentale degli Stati Uniti: circa il 15 per cento è assorbito dall'espansione della Basin and Range Province, circa il 70 per cento dallo scorrimento lungo la faglia di San Andreas e il rimanente 15 per cento si traduce in altre deformazioni nella regione costiera californiana. Come mostra il nostro diagramma vettoriale, non tutti i movimenti in queste tre zone di deformazione sono paralleli al movimento delle zolle. Sia l'espansione della Basin and Range Province, sia la deformazione nella regione costiera includono componenti significative perpendicolari al movimento delle zolle. Queste componenti si annullano quasi del tutto essendo pressoché uguali in modulo e opposte in verso. La compensazione geometrica fa pensare che sia l'espansione della Basin and Range Province a provocare la compressione nelle catene costiere della California.

Come può il nostro diagramma vettoriale aiutarci a comprendere la natura della deformazione che sta avendo luogo lungo la costa californiana? Dopo tutto, un vettore descrive solamente il movimento di insieme associato a una deformazione; fornisce pochissime informazioni sui dettagli della deformazione stessa. Un modo per analizzare il movimento stimato è quello di scomporre il vettore nelle sue due componenti: una componente di taglio, grosso modo parallela all'andamento della faglia di San Andreas, e una componente compressiva, perpendicolare alla faglia. Valutiamo che il tasso del movimento di taglio sia di circa sei millimetri all'anno e quello di compressione di circa otto millimetri all'anno.

La presenza di entrambe le componenti è in accordo con quanto si sa circa la tettonica della California, che ha molte faglie attive oltre a quella di San Andreas. L'esistenza di una componente compressiva è indicata dai piegamenti e

Il diagramma vettoriale indica come il movimento relativo delle zolle del Pacifico e nordamericana si suddivida tra le zone di deformazione negli Stati Uniti occidentali; si vedono il tasso relativo e la direzione del moto delle zolle (*in viola*), l'espansione della Basin and Range Province (*in blu*), lo scorrimento della faglia di San Andreas (*in arancione*) e la deformazione lungo la costa (*in rosso*).

Una nuova rete geodetica permette di seguire la deformazione a ovest della faglia di San Andreas. Le posizioni delle stazioni possono essere determinate con la precisione di circa un centimetro per mezzo di segnali inviati dai satelliti del Global Positioning System del Department of Defense. La rete consente di osservare cambiamenti nella posizione delle stazioni con il deformarsi della crosta. È stata messa in opera nel 1985 dal California Institute of Technology, il Massachusetts Institute of Technology, l'Università della California a Los Angeles e l'Università della California a San Diego, insieme a molti enti governativi.

dalle spinte che hanno luogo sulle aspre catene costiere. Gli effetti di questa componente possono pure essere rintracciati a mare, dove l'esplorazione sismica per la ricerca petrolifera ha rivelato un gran numero di pieghe da compressione nei sedimenti che si vanno rapidamente accumulando sulla piattaforma continentale. Inoltre, i dati sismologici e morfologici suggeriscono che parte della compressione venga esercitata in realtà sulle strutture appena a est della faglia di San Andreas; per esempio, il terremoto di Coalinga del 23 maggio 1983 ebbe il suo epicentro in un punto situato circa 27 chilometri a est della faglia di San Andreas. Sembra quindi probabile che il movimento di compressione comporti deformazioni distribuite su una regione che si estende dal margine occidentale della Great Valley alla piattaforma continentale, regione ampia circa 150 chilometri.

Il movimento di taglio, d'altro canto, può essere molto più localizzato. Gran parte di esso può essere attribuita, per esempio, al sistema di faglie San Gregorio-Hosgri, che decorre lungo la linea di costa da San Francisco alla base aeronautica di Vandenberg. Questo sistema di faglie è in gran parte sommerso, ma quando è visibile sulla terraferma mostra chiaramente che negli ultimi milioni di anni vi sono avvenute dislocazioni significative. Dai dati geologici si può desumere un tasso medio di movimento lungo la faglia anche di 13 millimetri all'anno. La sismicità di questo sistema di faglie, secondo le testimonianze storiche, è molto bassa; se il tasso di movimento si è mantenuto veramente sui 13 millimetri all'anno, il movimento di taglio si sarebbe dovuto esprimere in terremoti intensi, per quanto non frequenti, il che non è storicamente documentato. Non è chiaro se il tasso, più modesto, di sei millimetri all'anno dedotto dai nostri calcoli vettoriali avrebbe comportato eventi di questo genere. La questione non riveste comunque un interesse puramente accademico: la centrale nucleare di Diablo Canyon e la base di Vandenberg, attrezzata per il lancio dello Space Shuttle, si trovano entrambe nei pressi della faglia di Hosgri, come pure molti grandi centri urbani.

Queste ipotesi riguardanti la deformazione della California sono in gran parte indirette, comprendendo dati raccolti ben al di fuori della regione. Dato che dobbiamo includere assunti a priori ed effettuare numerose semplificazioni al fine di valutare il tasso e la direzione della deformazione, i nostri risultati presentano un ampio margine di incertezza e potrebbero essere falsati in partenza. Volendo giungere a valutazioni più accurate, dovremmo poter disporre di osservazioni dirette del movimento relativo di sezioni crostali su lati opposti di particolari zone di faglia.

Per rendere possibili tali osservazioni, un consorzio di quattro università sta collaborando con numerosi enti governativi per definire e controllare una rete geodetica che sarà disposta lungo la costa californiana da Monterey a San Diego e comprenderà la maggior parte delle isole al largo della California meridionale. La rete usufruisce di una tecnica di geodesia spaziale sviluppata recentemente, che si fonda su una «costellazione» di satelliti artificiali, chiamata Global Positioning System, messa in orbita e gestita dal Department of Defense.

Il sistema è destinato a localizzare navi, aerei e veicoli terrestri in tempo reale con una precisione di alcuni metri, ma la metodologia mutuata in parte dall'interferometria permette misurazioni tra stazioni fisse a terra con una precisione di pochissimi centimetri, su linee di base anche di molte centinaia di chilometri. Quindi il sistema ha circa la stessa precisione dell'interferometria, ma non richiede stazioni a terra grandi e costose come i radiotelescopi della VLBI. Rilevamenti ripetuti nei prossimi anni dovrebbero consentire ai geofisici di sviluppare una descrizione molto più dettagliata e accurata delle deformazioni nella regione costiera californiana. Queste descrizioni a loro volta renderanno possibili valutazioni più realistiche del rischio di terremoti in questa regione sismicamente attiva e densamente popolata e ci aiuteranno a comprendere in maniera molto più dettagliata una delle componenti della deformazione in atto negli Stati Uniti occidentali.

BIBLIOGRAFIA

HILL DAVID P., *Contemporary Block Tectonics: California and Nevada* in «Journal of Geophysical Research», 87, n. B7, 10 luglio 1982.

Special Issue on Satellite Geodynamics in «IEEE Transactions on Geoscience and Remote Sensing», GE-23, n. 4, luglio 1985.

MINSTER J. BERNARD e JORDAN THOMAS H., *Vector Constraints on Western U.S. Deformation from Space Geodesy, Neotectonics, and Plate Motions* in «Journal of Geophysical Research», 92, n. B6, 10 maggio 1987.

Farmaci attivati dalla luce

La fotoferesi, una nuova metodica terapeutica resa possibile grazie alla messa a punto di farmaci derivati dagli psoraleni, apre nuove prospettive nei confronti di alcuni tumori e malattie autoimmunitarie

di Richard L. Edelson

Uno dei più importanti scopi della medicina è lo sviluppo di mezzi terapeutici dotati di una precisa specificità: agenti che abbiano effetti solo sul tessuto bersaglio e su nessun altro. Molti antibiotici si avvicinano a questo comportamento ideale. A causa delle differenze esistenti fra le cellule batteriche e quelle umane, gli antibiotici possono sopprimere l'invasore e lasciare intatti i tessuti corporei. Nel cancro e nelle malattie autoimmunitarie (nelle quali il sistema immunitario aggredisce erroneamente i tessuti normali), tuttavia, è molto più difficile ottenere questa capacità discriminante. In entrambi i tipi di patologia, all'origine del danno tissutale non ci sono batteri o virus, ma le cellule stesse del corpo che assumono un comportamento aberrante. Chiaramente, distinguere le cellule sane da quelle malate non è un compito semplice. Per questo vi sono poche terapie efficaci per le malattie autoimmunitarie e quelle esistenti per il cancro spesso comportano gravi effetti collaterali.

Una strategia che ha cominciato a dare i suoi frutti è quella che utilizza farmaci attivati dalla luce. Poiché questi farmaci sono inerti quando non sono esposti a radiazioni della corretta lunghezza d'onda, essi permettono al medico di colpire solo i tessuti che sono esposti sia al farmaco sia alla luce. In particolare, i miei colleghi e io abbiamo usato un farmaco conosciuto come 8-MOP, la cui storia risale all'antico Egitto, per la cura del linfoma linfocitico a cellule *T* della cute o, più brevemente, linfoma cutaneo a cellule *T* (CTCL, da *cutaneous T-cell lymphoma*). Questa patologia è dovuta a una trasformazione maligna dei linfociti e ha prognosi infausta. Tuttavia, somministrando ai pazienti 8-MOP ed esponendo poi i loro linfociti a radiazione della lunghezza d'onda appropriata, abbiamo ottenuto risultati entusiasmanti. Questo metodo per il trattamento del CTCL, ora approvato dalla Food and Drug Administration degli Stati Uniti, potrebbe essere il precursore di un'ampia gamma di terapie basate su farmaci attivati dalla luce.

Il nostro lavoro ha finito per riunire due serie di esperimenti che all'inizio erano del tutto indipendenti: una sui tumori delle cellule del sangue e l'altra (portata avanti da Irun R. Cohen e colleghi del Weizmann Institute of Science in Israele) sulle malattie autoimmunitarie. Questi campi di ricerca avevano parecchi elementi in comune. In primo luogo sia le forme di tumore studiate da me e dai miei colleghi, sia le malattie autoimmunitarie studiate da Cohen derivano da anomalie dei linfociti *T*, globuli bianchi che hanno un ruolo fondamentale nell'organizzazione della risposta immunitaria. (È una particolare serie di cellule *T* quella che viene colpita nell'AIDS.)

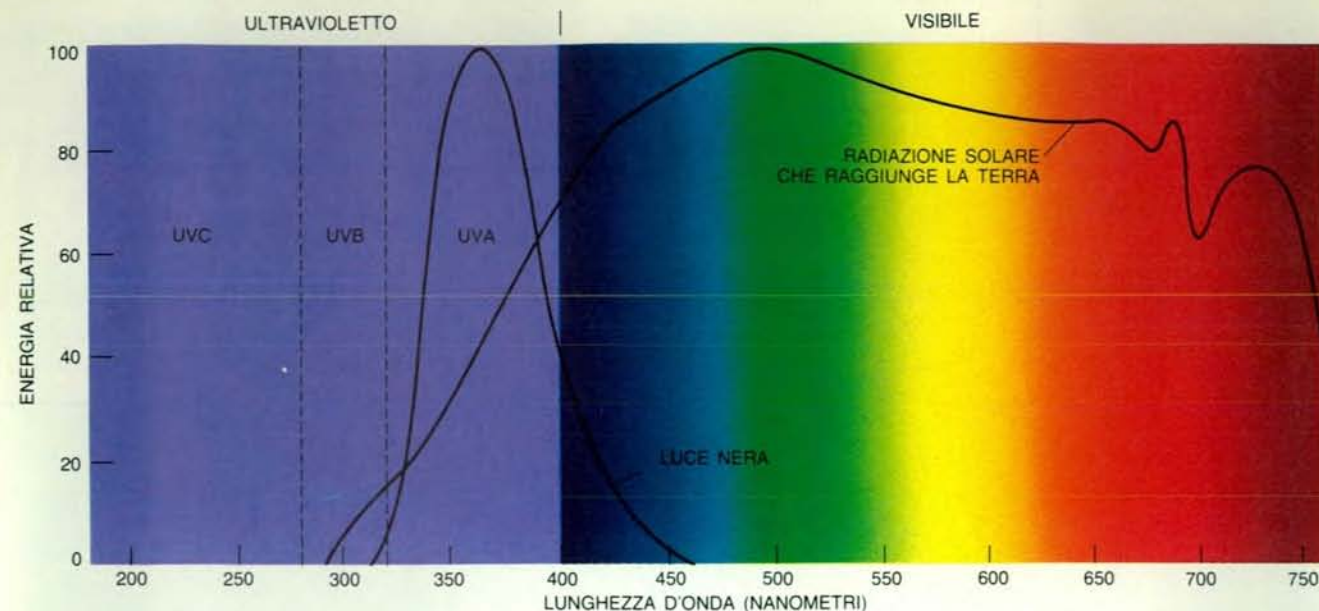
Di solito i medici che tentano di curare i due tipi di malattia incontrano problemi analoghi. La chemioterapia contro il cancro è di solito mirata contro cellule che si moltiplicano rapidamente. La tolleranza del paziente per questo tipo di terapia è limitata dalla sua tossicità per tessuti normali, come la mucosa intestinale, il midollo osseo e i bulbi capilliferi, che possiedono cellule che si dividono rapidamente. I soggetti in trattamento chemioterapico per un tumore spesso sono affetti da anemia (come conseguenza di una ridotta produzione di globuli rossi da parte del midollo osseo), emorragie (a causa di una diminuita produzione di piastrine, che sono necessarie per la coagulazione) e infezioni (per una ridotta produzione di globuli bianchi), oltre che da sanguinamento intestinale e caduta dei capelli.

Così come gli oncologi con la chemioterapia cercano di agire selettivamente sulle cellule maligne, gli immunologi clinici tentano di bloccare la crescita di cellule del sistema immunitario che reagiscono in modo anormale. In molti casi le cellule anormali sono linfociti *T*. Per esempio, nell'artrite reumatoide, una patologia aggressiva e invalidante che colpisce principalmente giovani donne, i linfociti attaccano i tessuti articolari come se quei normali tessuti corporei fossero invasori patogeni. I derivati cortisonici sono spesso usati per sopprimere i linfociti distruttivi. Tuttavia tra i comuni effetti collaterali dei derivati cortisonici vi sono ipertensione, diabete, cataratta, degenerazione ossea e una maggior facilità di contrarre infezioni che sarebbero facilmente vinte da una persona in buona salute.

Sebbene oggi sia chiaro che esiste una interessante sovrapposizione tra i problemi degli immunologi e quelli degli oncologi, all'inizio i miei colleghi e io ci siamo principalmente occupati di luce e cancro. Il ruolo della luce nel trattamento di alcune malattie è stato da tempo riconosciuto. Nel 1903 Niels R. Finsen ricevette il premio Nobel per la scoperta che le lesioni cutanee dovute alla tubercolosi spesso guarivano dopo esposizione alla radiazione ultravioletta. La tubercolosi è attualmente curata con gli antibiotici, ma la terapia con la luce è fondamentale nella cura di diverse malattie comuni, fra cui la psoriasi, l'acne e alcune forme di ittero nei neonati.

Molto prima di Finsen, tuttavia, gli antichi Egizi avevano osservato che un vegetale molto comune aveva proprietà medicamentose indotte dalla luce. La pianta, oggi classificata come *Ammi majus*, è un'erba che cresce sulle rive del Nilo. I medici del tempo avevano notato che chi ingeriva l'erba diventava insolitamente suscettibile all'abbronzatura. Essi fecero uso di questa proprietà per il trattamento della vitiligine, una malattia cutanea in cui la pelle presenta chiazze chiare più o meno estese perché alcune zone perdono la pigmentazione. Oggi si sa che i composti attivi di *Ammi* sono gli psoraleni, la classe di composti a cui appartiene l'8-MOP (la sua denominazione tecnica è 8-metossipsoralene). L'8-MOP, che è sia un farmaco anticancro, sia un regolatore immunitario, è un pro-

Prima del trattamento terapeutico le mani di un paziente affetto da linfoma cutaneo a cellule *T* (CTCL), una forma tumorale indotta dai linfociti *T*, appaiono gonfie e arrossate. La sintomatologia è dovuta all'abnorme proliferazione di linfociti *T* presenti nella cute. Nello stadio finale della malattia, che è resistente alle normali terapie, le cellule maligne invadono gli organi interni, distruggendoli.

Le mani dello stesso paziente sono state fotografate dopo il trattamento con un farmaco che viene attivato dalla luce, l'8-MOP. Il farmaco è inerte fino a quando non è esposto a luce ultravioletta di una particolare lunghezza d'onda. Allora si lega al DNA, danneggiando le cellule che si moltiplicano rapidamente, come i linfociti *T* maligni. La metodica, ideata dall'autore, è chiamata fotoferesi.

Lo spettro elettromagnetico si estende oltre il campo della radiazione visibile e comprende in particolare la radiazione ultravioletta. L'ultravioletto B (UVB) provoca le scottature. Nell'ultravioletto A (UVA) sono comprese le lunghezze d'onda che attivano l'8-MOP.

totipo per lo sviluppo di farmaci che possono essere attivati dalla luce.

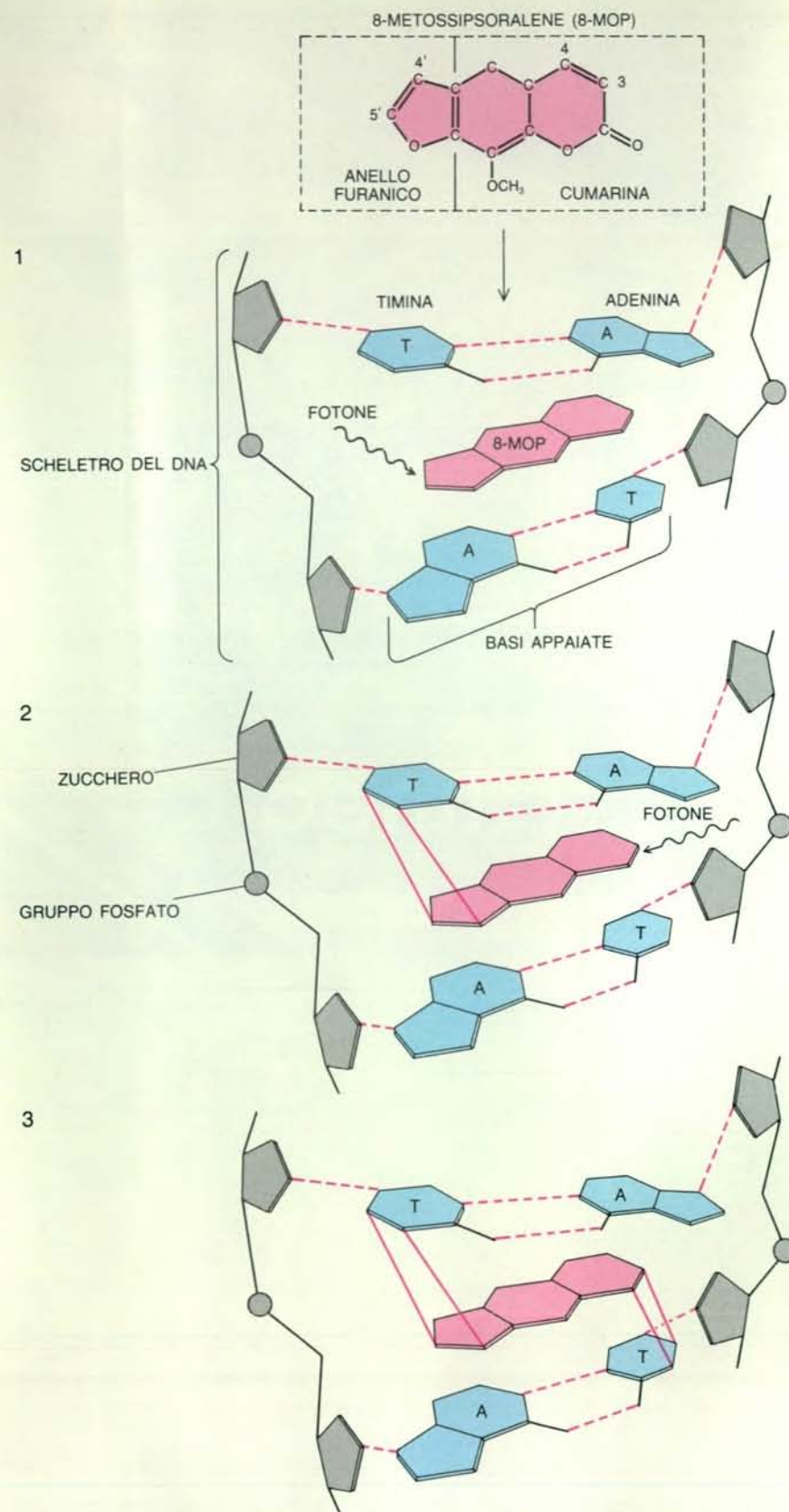
Gli psoraleni sono uno straordinario gruppo di composti ed è a loro che è stata dedicata in particolare la nostra attenzione. Gli psoraleni si trovano non soltanto in *Ammi*, ma anche (in minor quantità) nei fichi, nelle limette, nelle radici di pastinaca e in molti altri frutti e verdure. Una volta ingeriti con il cibo, gli psoraleni sono assorbiti nel tubo digerente e raggiungono i massimi livelli ematici e tissutali in un tempo variabile da una a quattro ore; nel giro di 24 ore essi sono escreti quasi completamente. La caratteristica più evidente di queste sostanze è la loro capacità di assorbire la luce ultravioletta e di essere attivate da essa. Prima dell'esposizione a questa radiazione le molecole sono inerti, ma in seguito esse si legano al DNA in modo tale che i due filamenti dell'elica rimangono uniti strettamente. Dal momento che la separazione dei filamenti è necessaria per la replicazione del DNA, gli psoraleni producono gravi danni alle cellule che si dividono rapidamente.

Sono molte le caratteristiche che rendono gli psoraleni particolarmente adatti a usi terapeutici. In assenza di luce essi sono assolutamente innocui, ma dopo l'attivazione interrompono efficacemente la funzione del DNA. Inoltre il tipo di radiazione che li attiva, la radiazione ultravioletta A, o UVA, passa facilmente attraverso il vetro e alcune materie plastiche, mentre la radiazione ultravioletta B (che causa le scottature) è filtrata da questi materiali. Questo fatto risulta estremamente utile nella progettazione delle apparecchiature. Cosa forse ancora più importante, gli psoraleni, una volta esposti alla radiazione ultravioletta, rimangono attivi solo per pochi milionesimi di secondo, un tempo abbastanza lungo per consentire una reazione chimica, ma sufficientemente breve per far sì che, al cessare dell'irradiazione, il farmaco ritorni immediatamente alla sua forma inerte. In sintesi, gli psoraleni, in quanto farmaci efficaci solo quando vengono sottoposti a radiazione di un tipo ben determinato, offrono al clinico una forma di terapia estremamente specifica.

È interessante notare che i moderni usi medici degli psoraleni in un certo senso sono la continuazione del loro utilizzo terapeutico nell'antico Egitto. Le prime ricerche moderne sull'8-MOP furono eseguite da Abdel M. El Mofty dell'Università del Cairo negli anni quaranta. El Mofty confermò quello che i suoi antichi predecessori già sapevano: l'ingestione di un estratto di *Ammi* associata a un'esposizione alla luce solare era un trattamento efficace contro la vitiligine. Le sue conclusioni suscitarono il profondo interesse di un gruppo di ricercatori della School of Medicine dell'Università del Michigan guidato da Aaron B. Lerner e Thomas B. Fitzpatrick, che effettuò i primi studi sull'8-MOP purificato. Questo lavoro, svolto nei primi anni cinquanta, dimostrò che il farmaco era notevolmente sicuro: il rapporto tra la dose massima tollerata e quella richiesta per un effetto clinico era molto alto.

Dopo quella dimostrazione iniziale, l'8-MOP rimase oggetto di un certo interesse sperimentale. Durante gli anni settanta, diversi ricercatori dimostrarono che il farmaco, in combinazione con la luce, poteva essere usato nel trattamento della psoriasi. (La psoriasi è una patologia non cancerosa in cui le cellule dell'epidermide si dividono molto più rapidamente della norma, dando come risultato una cute spessa e a scaglie). Circa nello stesso periodo, Barbara Gilchrest e colleghi della Harvard Medical School scoprirono che le lesioni epidermiche localizzate prodotte dal linfoma cutaneo a cellule *T* scomparivano in risposta a un trattamento analogo.

Questa scoperta fu particolarmente interessante, sia perché questa degenerazione maligna dei linfociti *T* è la più comune negli adulti, sia perché era particolarmente resistente alla maggior parte delle terapie note negli anni settanta. Nelle sue fasi iniziali la malattia è caratterizzata da lesioni cutanee contenenti molte cellule *T* maligne. Potrebbe sembrare strana la presenza di globuli bianchi nella cute, ma di fatto essi sono molto comuni poiché la cute svolge importanti funzioni di tipo immunitario (si veda l'articolo *La funzione immunologica della pelle* di Richard L. Edelson e Joseph M. Fink in «Le Scienze» n. 204, agosto 1985). Nella fase finale della malattia le cellule maligne si diffondono in tutto il corpo e la prognosi è infausta. Infatti la sopravvivenza media di un paziente non curato è di soli cinque anni dal momento della diagnosi confermata da un esame bioptico.

Sembrava possibile che la combinazione sperimentale di somministrazione di 8-MOP ed esposizione di tutto il corpo a una intensa radiazione UVA compiuta dalla Gilchrest funzionasse poiché il farmaco e la luce associati danneggiavano direttamente le cellule *T* maligne nelle lesioni cutanee. Si sapeva da tempo che il 25 per cento del flusso ematico passa attraverso i vasi cutanei e che una parte

L'8-MOP si lega al DNA dopo essere stato attivato dalla radiazione ultravioletta. L'8-MOP (8-metossipsoralene), mostrato in alto, è un composto semplice costituito da due strutture: un anello furanico e una cumarina. Dopo aver raggiunto il nucleo di una cellula, si inserisce tra le basi appaiate della catena di DNA (*1*). Assorbito un fotone UVA, la molecola forma una coppia di legami con una base nucleotidica su un filamento del DNA (*2*). Dopo aver assorbito un secondo fotone, l'8-MOP può legarsi a una base sull'altro filamento del DNA (*3*). L'unione dei due filamenti dell'elica impedisce la replicazione del DNA.

della radiazione UVA può attraversare lo spessore della cute. Oltre a ciò il lavoro svolto nel mio laboratorio e in quelli di Kenneth H. Kraemer al National Cancer Institute e di Warwick L. Morison a Harvard ha dimostrato che i linfociti *T* sono assai sensibili agli effetti dannosi dell'8-MOP e della radiazione ultravioletta. Queste scoperte suggerivano la fantastica possibilità di rimuovere le cellule maligne dall'organismo per sottoporle direttamente all'azione tossica combinata del farmaco e della luce.

L'idea di questa soluzione sorse in me spontaneamente, anche perché stavo già lavorando a un sistema per estrarre il sangue dal corpo e rimuovere da esso le cellule maligne. Il sistema, chiamato leucaferesi, prevede la separazione dei vari componenti del sangue tramite centrifugazione. Poiché i componenti hanno densità diverse, essi si segregano nella provetta di centrifugazione e possono essere separati. La separazione, però, non è estremamente selettiva e di conseguenza alcuni globuli rossi, piastrine e globuli bianchi normali sono rimossi insieme ai linfociti alterati. Quando applicammo incondizionatamente il metodo della leucaferesi nel tentativo di rimuovere la maggior parte delle cellule maligne, i pazienti divennero anemici (a causa della perdita di globuli rossi) e vulnerabili alle infezioni (per la perdita di globuli bianchi).

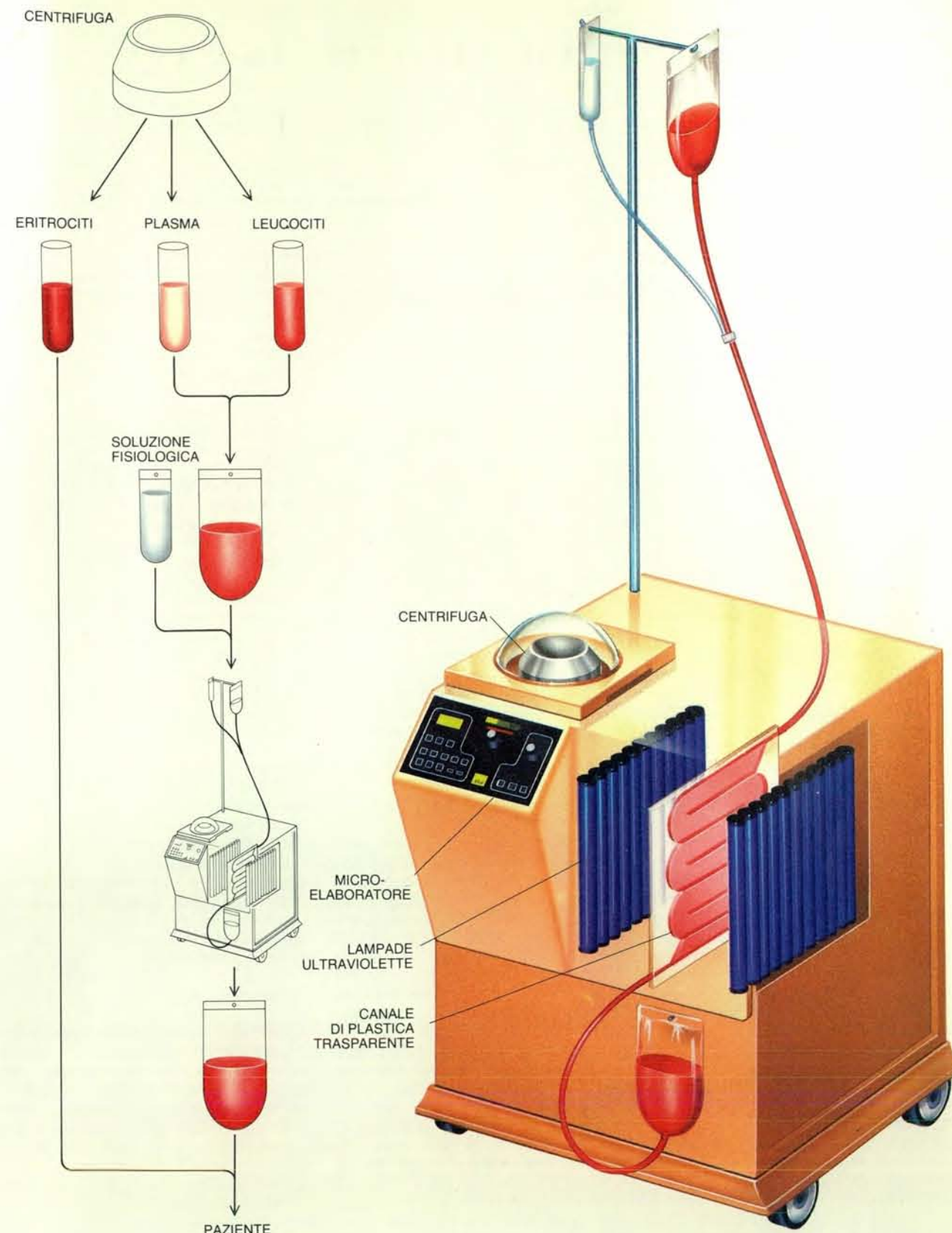
Sembrava che l'efficacia di questo metodo potesse essere aumentata se i globuli bianchi venivano esposti all'8-MOP e alla luce all'esterno dell'organismo. Dal momento che sia i globuli rossi sia le piastrine non sono dotati di nucleo, soltanto i globuli bianchi sarebbero stati sensibili al farmaco che ha la proprietà di legarsi al DNA. Inoltre, una volta rientrato nell'organismo, il farmaco sarebbe ritornato inerte e nessun tessuto oltre al sangue sarebbe stato esposto all'8-MOP attivo. I miei colleghi e io proseguimmo gli studi necessari per sperimentare questo nuovo, promettente metodo, anche se eravamo alquanto preoccupati per le conseguenze che potevano derivare dalla reimmissione di un gran numero di linfociti *T* danneggiati nel flusso sanguigno del paziente. Quello che non potevamo prevedere era che questo fatto si sarebbe rivelato l'elemento fondamentale della terapia.

Dopo aver condotto ricerche preliminari intese a stabilire il dosaggio ottimale del farmaco e dell'irradiazione, eravamo pronti a procedere. La radiazione ultravioletta A attraversa la plastica acrilica trasparente in modo assai soddisfacente, sicché questo materiale fu scelto per la realizzazione del sistema di esposizione, un progetto al quale i miei colleghi e io lavorammo in collaborazione con i tecnici della Therakos Inc., una consociata della Johnson & Johnson. I raggi UVA sono un tipo di radiazione ultravioletta debole e, sebbene questa loro proprietà

sia utile per minimizzare gli effetti collaterali dovuti all'esposizione, essa implica che il lume del canale attraverso cui viene fatto scorrere il sangue per essere esposto alla radiazione debba essere estremamente ridotto. Se così non fosse, la radiazione verrebbe per la maggior parte assorbita dai globuli rossi prima di poter raggiungere i leucociti. Infatti, negli apparecchi di terza generazione deputati a questo procedimento, che è stato denominato fotoferesi, uno strato di sangue dello spessore di un solo millimetro viene fatto passare attraverso due sorgenti di radiazione ultravioletta ad alta intensità.

Nel corso del trattamento circa 500 millilitri di sangue precedentemente separato nei suoi componenti (circa lo stesso volume che viene estratto in una donazione di sangue) passano davanti alla sorgente luminosa. Dal lavoro di dosaggio già effettuato conoscevamo la quantità di luce e di farmaco a cui doveva in media essere esposto un linfocita *T* per venire inattivato. Per stabilire la durata dell'esposizione del sangue necessaria per ottenere quei livelli, Francis P. Gasparro, Regina M. Santella e io preparammo anticorpi monoclonali per il complesso costituito da 8-MOP e da DNA. Una preparazione di anticorpi monoclonali si lega specificamente a un solo tipo di molecola. Utilizzando una preparazione di questo tipo fummo in grado di stabilire la durata dell'esposizione necessaria per ottenere il numero richiesto di legami tra 8-MOP e DNA in ciascuna cellula *T*. I nostri risultati dimostravano che un'esposizione di 150 minuti alla radiazione ultravioletta nell'apparecchio per la fotoferesi sarebbe stata sufficiente.

Poiché la fotoferesi rappresentava una forma di terapia sperimentale e potenzialmente pericolosa, il gruppo iniziale di pazienti doveva rispondere a criteri di scelta particolarmente severi. La patologia scelta per il trattamento era la forma leucemica del linfoma cutaneo a cellule *T*. La malattia è provocata da una massiccia proliferazione di un singolo clone di cellule *T*. Esistono milioni di differenti cloni (popolazioni geneticamente identiche) di cellule *T* nel sistema immunitario. Le cellule di ciascun clone possiedono un recettore in grado di riconoscere un'unica molecola estranea. Quando quella molecola è presente nel corso di una malattia, il clone corrispondente di cellule *T* si moltiplica come parte della risposta immunitaria. In caso di trasformazione maligna, tuttavia, il clone aberrante prolifera massicciamente, diventando predominante nella popolazione dei globuli bianchi e, alla fine, uccidendo il paziente.

Nella forma leucemica del linfoma cutaneo a cellule *T* i componenti del clone aberrante migrano (attraverso i vasi sanguigni e linfatici) fra la cute, che è infiltrata a tal punto che l'intera superficie

La fotoferesi comporta il prelievo di 500 millilitri di sangue (circa lo stesso quantitativo estratto in una donazione di sangue). L'apparecchio per la fotoferesi, mostrato a destra, è stato messo a punto dai tecnici della Therakos, Inc., (una consociata della Johnson & Johnson) in collaborazione con l'autore. Dopo aver separato i componenti del sangue mediante centrifugazione, il plasma (la frazione liquida) e i leucociti (i globuli bianchi), vengono uniti a una soluzione fisiologica. Nella macchina un velo sottile della sospensione contenente i leucociti viene fatto scorrere in un canale di plastica trasparente collocato fra due serie di lampade UVA ad alta intensità. Dopo l'irradiazione, i globuli rossi sono ricombinati con gli altri componenti del sangue che viene poi trasfuso nel paziente.

corporea appare arrossata e gonfia, e gli altri tessuti corporei. Questa patologia è estremamente debilitante e, alla fine, letale. Quando i pazienti affetti dalla forma leucemica della malattia sono trattati con le terapie normali, essi sopravvivono in media per meno di tre anni e muoiono a causa di infezioni opportuniste o per la distruzione di organi vitali da parte delle cellule maligne. Gli studi effettuati dal mio gruppo e da altri hanno dimostrato che una leucaferesi intensa (due o tre volte alla settimana per periodi prolungati) poteva indurre miglioramenti temporanei. Tuttavia il trattamento era costoso e il miglioramento delle condizioni dei pazienti di breve durata. Nel 1982, quando iniziammo i nostri tentativi con la fotoferesi, nessuno dei principali ospedali faceva più un uso regolare di quella tecnica per il trattamento del linfoma cutaneo a cellule *T*.

Iniziammo i trattamenti secondo uno schema prudente, volto soltanto a stabilire la tossicità del nuovo metodo. A intervalli di un mese, i pazienti venivano sottoposti a fotoferesi in due giorni successivi. Poiché un numero estremamente grande di cellule maligne si era già accumulato nella cute, nelle ghiandole linfatiche e nella milza, un trattamento così poco frequente del sangue avrebbe reso impossibile l'irradiazione di più del 10 per cento di esse. Perciò prevedemmo che, dopo aver verificato la sicurezza del metodo, la frequenza avrebbe dovuto essere di gran lunga aumentata. Si immagini la nostra sorpresa quando quattro dei primi cinque pazienti risposero positivamente dopo 6-10 trattamenti soltanto e quando, alla fine, la cute gravemente compromessa di due di loro non presentò più lesioni.

Questo era un mistero che richiedeva una spiegazione. Una cosa sembrava chiara fin dall'inizio: la maggior efficacia della fotoferesi rispetto alla leucaferesi stava nel fatto che, dopo essere stati danneggiati, i globuli bianchi, anziché essere eliminati, venivano reimmessi nel circolo del paziente. Questo aspetto della fotoferesi, che aveva inizialmente sollevato le maggiori preoccupazioni, sembrava essere l'elemento determinante della sua efficacia. Ma perché?

Nessuna risposta era immediatamente disponibile, ma nel frattempo un tentativo di maggiori proporzioni fu organizzato in collaborazione con la Food and Drug Administration. Un totale di 37 pazienti che non rispondevano alle consuete terapie fu ricoverato negli istituti medici della Yale University, della Columbia University, dell'Università della Pennsylvania, dell'Università della California a San Francisco, dell'Università di Vienna e dell'Università di Düsseldorf. Di questi, 27 risposero positivamente al trattamento e fra di essi 20 appartenevano al gruppo di 26 pazienti che presentavano un coinvolgimento dei linfonodi (che indica uno stadio avanzato della malattia). Non soltanto tale livello di risposta è stato notevole per una malattia resistente alle consuete terapie, ma anche gli effetti collaterali del trattamento sono stati minimi; la fotoferesi è ora consentita dalla Food and Drug Administration come normale forma di terapia per lo stadio avanzato del linfoma cutaneo a cellule *T*.

Per i miei collaboratori e per me, nel nostro ruolo di medici, questo era un risultato soddisfacente. Tuttavia come ricercatori continuavamo a essere tormentati dal problema di stabilire perché il trattamento fosse così efficace. Forse la scoperta più importante fatta nella seconda serie di esperimenti fu che i pazienti che rispondevano meglio al trattamento erano quelli il cui sistema immunitario era inizialmente più efficiente. Un'analisi effettuata dalla mia collega Carole Berger dimostrò che tutti i pazienti che avevano risposto bene al trattamento presentavano una diminuzione stabile del numero di cellule *T* maligne. Le cellule *T* normali danneggiate dalla fotoferesi, invece, erano rapidamente rimpiazzate dalle riserve dell'organismo. In qualche modo la terapia luce-farmaco (fotochemioterapia) provocava una distruzione selettiva del clone maligno.

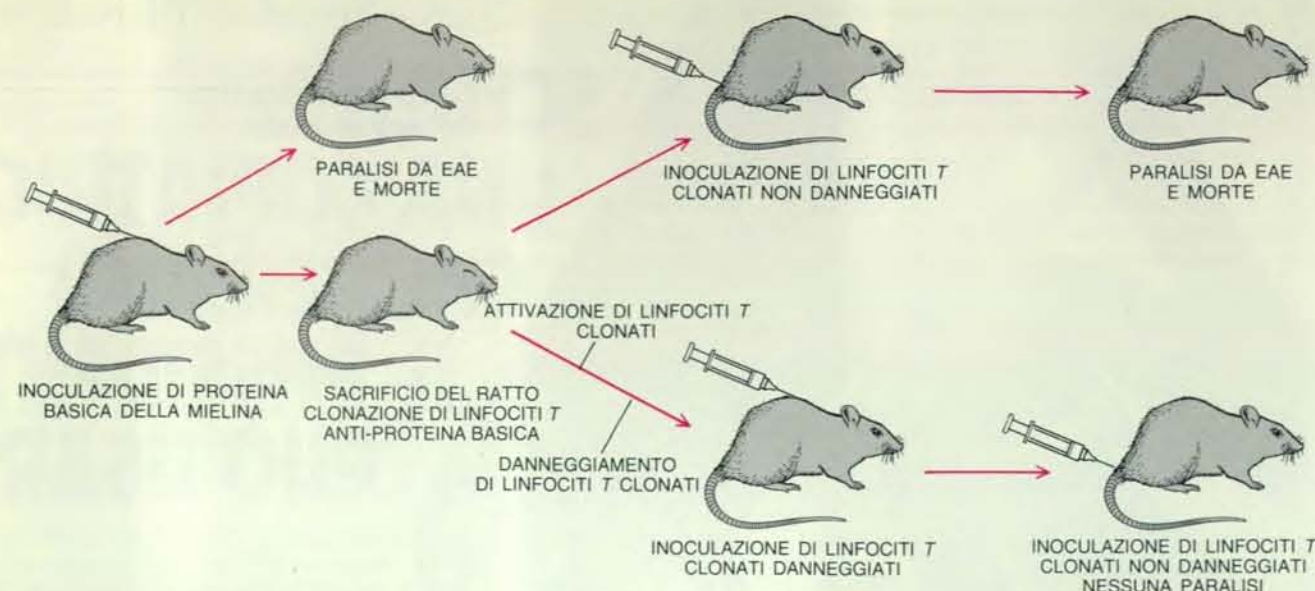
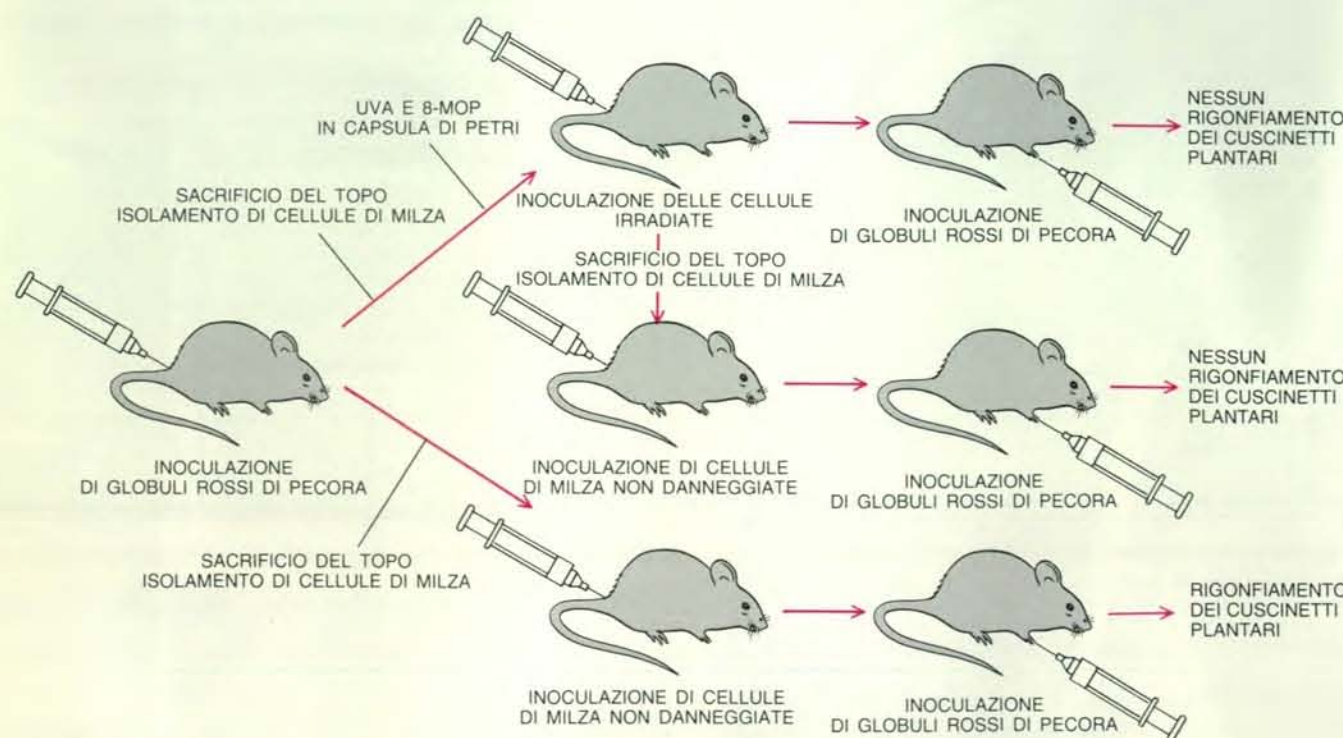
Tuttavia si trattava di un risultato paradossale, poiché il trattamento in se stesso era chiaramente aspecifico: tutti i globuli bianchi venivano in media esposti alla stessa quantità di 8-MOP e di energia radiante. La soluzione più probabile del paradosso era che doveva esistere un numero di componenti del clone maligno tanto maggiore rispetto a quello di ogni altro clone di cellule *T* che, una volta reimmesse nel corpo, le cellule malate danneggiate inducevano - a differenza delle altre - un'intensa risposta specifica. Spesso un clone maligno si moltiplica a tal punto che il numero dei suoi membri può superare quello di ogni altro singolo clone anche di un milione a uno. Quindi, quando questi linfociti vengono reimmessi in circolo, la milza organizza una potente risposta immunitaria contro di essi, ma non contro le cellule normali, che sono troppo poche per suscitare una risposta altrettanto energica.

Ma, ci si potrebbe chiedere, come può il sistema immunitario rispondere a un gruppo di cellule *T* dello stesso organismo a cui appartiene, maligne o no? La risposta ci conduce al cuore della moderna immunologia e, in particolare, alle ricerche di Irun Cohen (si veda l'articolo *Immunità e autoimmunità* di Irun R. Cohen in «Le Scienze» n. 238, giugno 1988). Il recettore che consente a una cellula *T* di riconoscere un antigene specifico, o molecola estranea, è esso stesso una proteina che può essere riconosciuta dal sistema immunitario, così come è riconosciuta la proteina di un agente patogeno. La parte caratteristica di un recettore che gli consente di adattarsi al suo antigene è chiamata idiotipo; l'adattamento tra lo stesso recettore e un secondo recettore presente su un altro linfocita e, a sua volta, in grado di combinarsi con il primo è chiamato interazione anti-idiotipica. Dato che il sistema immunitario nel suo complesso non viene danneggiato dalla fotoferesi, la spiegazione più verosimile del fenomeno da noi osservato è che il sistema immunitario (in particolare la milza) organizzi una potente risposta tramite cellule dotate di recettori anti-idiotipici che riconoscono il clone maligno.

Queste osservazioni suggerivano che i nostri risultati avrebbero potuto essere meglio interpretati alla luce degli studi condotti da Cohen e dal suo gruppo sugli animali. Essi si occupavano di parecchie malattie sperimentali nelle quali una risposta autoimmunitaria può essere provocata tramite l'inoculazione di una sostanza specifica in un animale da laboratorio. Alla base della malattia c'è il fatto che l'antigene inoculato assomiglia a molecole che si trovano nei tessuti corporei e, di conseguenza, le cellule *T*, oltre che l'antigene, attaccano anche il tessuto normale. Per esempio, i ratti inoculati con la proteina basica (un componente della mielina che costituisce la guaina di alcuni nervi) sviluppano una malattia che paralizza il sistema nervoso, chiamata encefalomielite autoimmune sperimentale (EAE).

La scoperta più sorprendente effettuata da Cohen fu che, se si poteva identificare il clone di cellule *T* responsabile della risposta autoimmunitaria, esso avrebbe potuto essere usato come «vaccino». Il clone anti-proteina basica, per esempio, veniva estratto da ratti che avevano sofferto di encefalomielite autoimmune sperimentale ed erano guariti. Le cellule venivano sottoposte agli effetti di particolari sostanze attivanti e i recettori sulla loro superficie venivano aggregati con vari sistemi per renderli più immunogenici. Quando le cellule trattate venivano reiniettate in altri ratti, quegli animali erano protetti contro l'insorgenza dell'encefalomielite: le cellule trattate agivano come «vaccino» contro l'autoimmunità.

Sembrava che quei risultati presentassero notevoli analogie con i nostri. Infatti, avevamo rimosso le cellule *T* dal corpo, le avevamo alterate e quindi le avevamo reintrodotte. Inoltre, molte delle cellule con cui avevamo lavorato appartenevano a un clone particolare, quello maligno. Forse la fotoferesi costituiva la «vaccinazione» dei pazienti affetti da linfoma cutaneo a cellule *T* contro il loro stesso cancro.

Per scoprire se questa ipotesi fosse esatta, diversi metodi sperimentali che simulavano la fotoferesi furono provati nel nostro laboratorio da Carole Berger, Maritza Perez e Liliane Laroche. In uno di questi tentativi un ceppo di topi geneticamente identici ottenuto con accoppiamenti tra consanguinei fu inoculato

Il «vaccino» contro l'autoimmunità, scoperto da Irun R. Cohen del Weizmann Institute of Science in Israele, opera con un meccanismo analogo a quello della fotoferesi. Ratti a cui sia stata inoculata la proteina basica della mielina sviluppano una malattia autoimmunitaria, l'encefalomielite autoimmune sperimentale (EAE), che rappresenta una reazione contro la mielina del proprio sistema nervoso. Questa reazione è mediata dai linfociti *T* che portano recettori per la proteina basica. A causa della malattia alcuni ratti muoiono, ma altri si riprendono. Per preparare il vaccino i linfociti *T* responsabili della risposta autoimmunitaria, dopo essere stati prelevati, sono clonati, attivati e danneggiati esponendoli a una pressione elevata o a opportune sostanze chimiche. Quando i linfociti *T* vengono inoculati in altri ratti, questi risultano immunizzati contro la malattia; se invece i linfociti *T* vengono iniettati senza attivazione e danneggiamento, provocano la malattia. Per quanto promettente, il metodo presenta controindicazioni per l'uso terapeutico su esseri umani. Nelle malattie dell'uomo è attualmente difficilissimo o impossibile isolare il corrispondente clone di linfociti *T*. Inoltre, le fasi di danneggiamento e di attivazione potrebbero provocare effetti collaterali dopo la reintroduzione delle cellule nel paziente.

Un esperimento condotto dall'autore dimostra che l'8-MOP e la radiazione UVA possono impedire una specifica risposta immunitaria con un meccanismo analogo a quello scoperto da I. R. Cohen. Topi inoculati con globuli rossi di pecora sviluppano una risposta immunitaria. I linfociti *T* responsabili della reazione, concentrati nella milza, possono essere rimossi e sottoposti all'azione congiunta di 8-MOP e UVA in una capsula da laboratorio. Quando vengono inoculati in altri topi, essi impediscono specificamente la risposta immunitaria ai globuli rossi di pecora. Contrariamente al metodo sviluppato da Cohen, non è necessaria né la clonazione né un'ulteriore attivazione. L'azione congiunta della luce e dell'8-MOP è alla base della fotoferesi, metodica che ha già applicazioni cliniche.

con globuli rossi di pecora, che producono una vivace risposta immunitaria. Alcune cellule della milza di questi topi furono prelevate ed esposte a 8-MOP e a radiazione ultravioletta in capsule da laboratorio. Le cellule furono poi iniettate in topi geneticamente identici. Quando gli animali così «vaccinati» furono esposti ai globuli rossi di pecora, essi non mostrarono alcuna risposta immunitaria. La ragione più probabile è che la «vaccinazione» avesse indotto una forte risposta immunitaria anti-idiotipica diretta proprio contro le cellule della milza che avrebbero dovuto servire da protezione contro le cellule estranee.

Questo risultato (e molti altri analoghi) indicò chiaramente che la fotoferesi poteva condurre agli effetti osservati da Cohen e colleghi. Infatti, con il senno di poi, comprendevamo ora che l'8-MOP e la radiazione ultravioletta A sono i mezzi ideali per ottenere questi effetti. La capacità di attivare un potente farmaco in modo che danneggi un gran numero di cellule *T* anormali, che portano un marcatore di superficie specifico, prima che queste vengano reimmesse in un sistema immunitario intatto rappresenta la soluzione più adatta per stimolare una forte risposta immunitaria specifica contro quelle cellule. Ora è possibile capire perché la fotoferesi è così efficace: le cellule danneggiate del clone maligno, infatti, inducono il sistema immunitario a distruggere specificamente quel clone, facendo in modo che l'organismo si difenda da solo dal tumore che l'ha colpito.

Un'importante questione teorica che deve ancora essere risolta è perché, se ci sono tante cellule del clone maligno nel sangue del paziente, il sistema immunitario non organizzi da sé la risposta difensiva. Dopo tutto, nella leucemia il clone maligno costituisce spesso più del 20 per cento di tutti i leucociti. Il lavoro di Cohen suggerisce che nel sistema immunitario di un animale le cellule *T* debbano prima essere attivate e poi danneggiate affinché possano suscitare la risposta desiderata. Le cellule *T* maligne sono già attivate (come parte del processo canceroso), ma sembra che la fotoferesi sia necessaria perché il danno inflitto alle cellule ne aumenti in qualche modo l'immunogenicità. Sono in corso studi nel nostro e in altri laboratori proprio per scoprire la ragione di questo fatto.

La fotoferesi potrà in futuro essere usata per il trattamento di una grande varietà di malattie che implicano una proliferazione di cloni specifici di cellule *T*. Quando queste applicazioni saranno possibili, tuttavia, è probabile che la tecnica sarà stata ampiamente modificata e migliorata. Uno di questi miglioramenti riguarderà il farmaco stesso. Un analogo di sintesi dell'8-MOP, l'amminometil-trimetilpsoralene, è più solubile in acqua della sostanza naturale e ha anche una maggior affinità per il DNA. Queste proprietà conducono, come hanno dimostrato Gasparro e colleghi, a un'attività per molecola molto maggiore. In futuro saranno certamente trovati o sintetizzati altri farmaci molto adatti alla fotoferesi.

Verranno inoltre messi a punto sistemi migliori e più specifici per portare il farmaco e la luce attivante sul tessuto bersaglio. Un sistema sperimentale per questo scopo è stato studiato nel nostro laboratorio da Shrishailam Yemul, Alison Estabrook, Hagen Bayley, Carole Berger e da me. In questo sistema gli anticorpi monoclonali sono posti sulla superficie di vescicole (piccole sacche) costituite dallo stesso tipo di doppio strato lipidico che compone la membrana cellulare esterna. Scegliendo l'anticorpo appropriato, siamo stati in grado di indurre le vescicole a legarsi preferenzialmente a particolari tipi di cellule. Molecole di pirene, una sostanza fotoattiva, sono incorporate nelle vescicole; quando queste raggiungono il bersaglio e il pirene viene attivato dalla luce, esso perfora la membrana esterna della cellula, provocandone la morte.

Molti altri sistemi sperimentali sono in fase di messa a punto, e tutti offrono i vantaggi teorici di una grande specificità. In questi sistemi il farmaco potrebbe essere attivato non soltanto all'esterno del corpo, come negli attuali apparati per la fotoferesi, ma anche al suo interno, a patto di trovare un metodo per portare la luce dove è necessaria. Le molecole ibride nei sistemi di questo tipo possono essere considerate i prototipi di farmaci appartenenti a una nuova categoria di agenti farmacologici: quelli attivati dalla luce. Le attuali conoscenze tecniche consentono di modificare quasi tutti i farmaci in modo tale che essi possano essere attivati solo dalla luce. Perciò potremmo un giorno ripensare agli psoraleni non modificati come a poco più che precursori di una categoria di farmaci completamente nuova.

BIBLIOGRAFIA

KNOBLER ROBERT M. ed EDELSON RICHARD L., *Cutaneous T Cell Lymphoma* in «The Medical Clinics of North America», 70, n. 1, gennaio 1986.

YEMUL SHRISHAILAM, BERGER CAROLE, ESTABROOK ALISON, SUAREZ SYLVIA, EDELSON RICHARD e BAYLEY HAGEN, *Selective Killing of T Lymphocytes by Phototoxic Liposomes* in «Proceedings of the National Academy of Sciences of the United States of America», 84, n. 1, gennaio 1987.

EDELSON RICHARD e altri *Treatment of Cutaneous T-Cell Lymphoma by Extracorporeal Photochemotherapy: Preliminary Results* in «The New England Journal of Medicine», 316, n. 6, 5 febbraio 1987.

Ombreggiatura e percezione delle forme

Il cervello interpreta l'ombreggiatura delle immagini bidimensionali per ricostruire la sensazione di uno spazio a tre dimensioni e per farlo presuppone automaticamente che la luce provenga da un'unica direzione

di Vilayanur S. Ramachandran

L'esperienza visiva della realtà che ci circonda si basa su immagini di tipo bidimensionale: configurazioni piane caratterizzate da una variabile intensità luminosa e da diversi colori, che cadono su un unico piano di cellule nella retina. Eppure noi, sulla base di tali immagini piane, riusciamo a percepire la solidità dei corpi e la loro profondità. La nostra percezione della tridimensionalità degli oggetti dipende da un certo numero di indizi sulla profondità inclusi nell'immagine che si forma sulla retina: ombreggiatura, prospettiva, occultamento parziale di un oggetto da parte di un altro e disparità stereoscopica. In un qualche modo misterioso il cervello è in grado di sfruttare questi indizi per risalire alle forme tridimensionali degli oggetti.

Fra i molti meccanismi usati dall'apparato visivo per percepire la terza dimensione, la capacità di sfruttare l'ombreggiatura è forse il più primitivo. Una ragione a sostegno di questa ipotesi sta nel fatto che in natura molti animali hanno evoluto parti ventrali più chiare presumibilmente per rendersi meno visibili ai predatori. Questa «ombreggiatura alla rovescia» compensa gli effetti di luci e ombre dovuti all'illuminazione solare dall'alto, e presenta almeno due vantaggi: riduce il contrasto con lo sfondo e «appiattisce» la forma percepita dell'animale. La prevalenza di questo tipo di colorazione in molte specie animali, fra cui molti pesci, suggerisce che l'ombreggiatura possa essere una fonte di informazione cruciale ai fini della percezione tridimensionale.

I pittori, ovviamente, da lungo tempo sfruttano la distribuzione di luci e ombre per rendere nelle loro immagini la vivida illusione della profondità. Gli psicologi, invece, non hanno dedicato molte ricerche alla scoperta dei meccanismi con cui l'occhio e il cervello utilizzano l'informazione contenuta nell'ombreggiatura. I miei colleghi e io abbiamo perciò intrapreso una serie di esperimenti per individuare dove risiedano alcuni di questi meccanismi.

Cominciammo le nostre ricerche generando, con l'aiuto di un calcolatore, una serie di visualizzazioni di oggetti semplici in cui l'impressione della profondità veniva resa solo in base a tenui variazioni nell'ombreggiatura. Ci assicurammo che le immagini fossero prive di oggetti o strutture complessi, poiché il nostro obiettivo era quello di separare i meccanismi cerebrali che elaborano l'informazione contenuta nell'ombreggiatura dai meccanismi di livello superiore, che pure possono contribuire alla percezione della profondità elaborando l'informazione visiva qual essa si presenta nella vita reale.

Per sviluppare i nostri esperimenti decidemmo di ricorrere a forme circolari ombreggiate, le quali danno luogo a una forte sensazione di profondità (*si veda l'illustrazione a pagina 74, in alto*). Le forme sembrano alternativamente sporgere verso l'esterno, come uova, o rientrare verso l'interno, come le cavità di un contenitore per uova. Esse risultano ambigue perché il cervello non può stabilire da quale direzione la luce stia provenendo. Con un po' di sforzo è possibile spostare mentalmente la sorgente di luce e invertire in tal modo la percezione del rilievo degli oggetti.

È degno di nota che quando si inverte mentalmente la percezione del rilievo di un oggetto, analogamente appaiono invertiti tutti gli altri oggetti presenti nella stessa visualizzazione. Questa circostanza solleva un problema interessante: la propensione a vedere tutti gli oggetti presenti nella visualizzazione come simultaneamente convessi o concavi si fonda su una tendenza a vederli tutti come dotati della stessa profondità oppure sul tacito assunto che esista un'unica sorgente di luce? Per accertarlo abbiamo creato una visualizzazione in cui gli elementi di una serie di oggetti sono le immagini speculari degli oggetti contenuti in una seconda serie (*b*). In questa visualizzazione, quando si vedono gli oggetti di una serie come se fossero convessi, quelli dell'altra serie sono sempre percepiti come concavi.

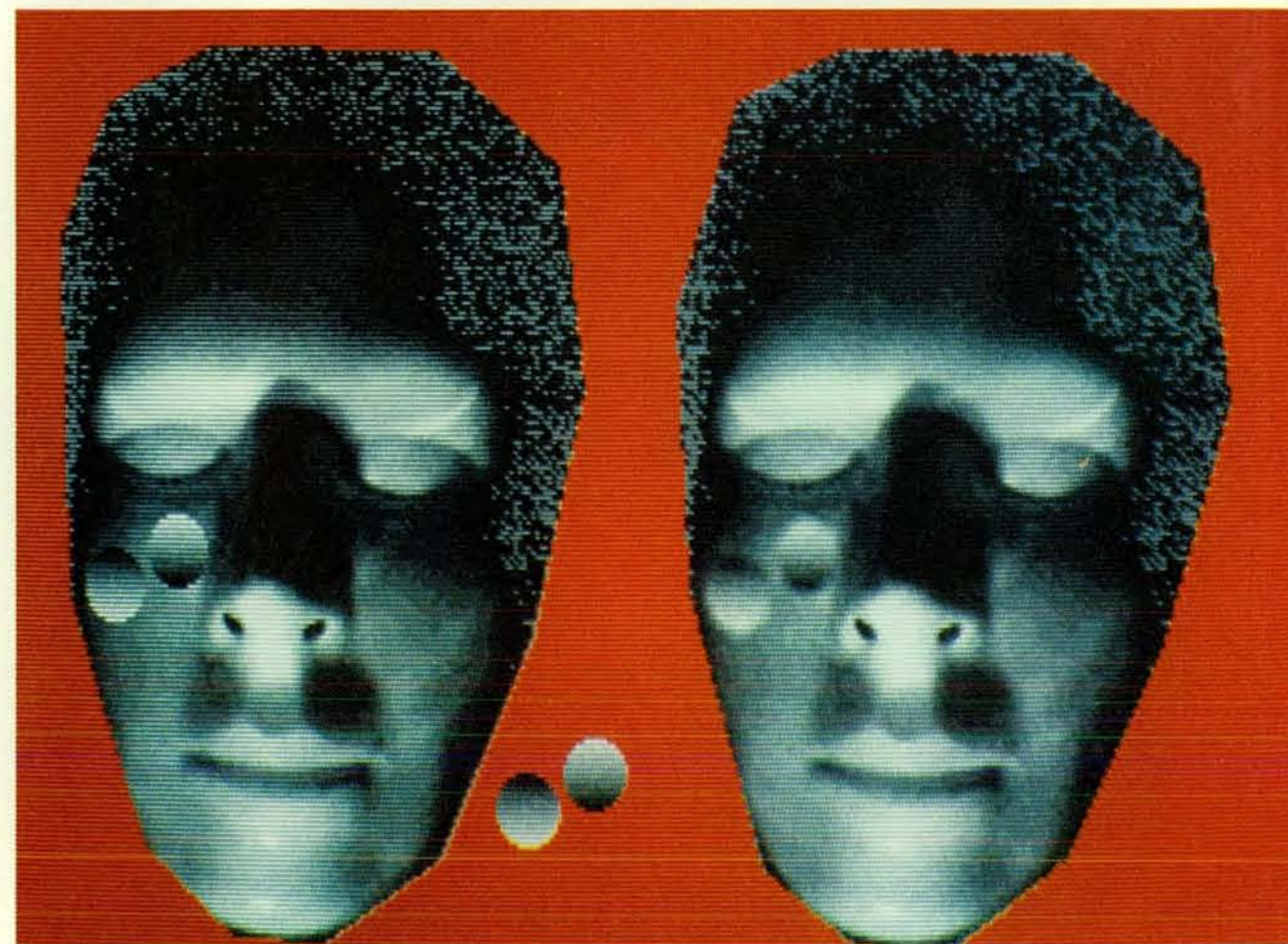
Grazie all'esecuzione di questo elementare esperimento siamo giunti a due conclusioni: innanzitutto, la derivazione della forma di un oggetto dalla sua ombreggiatura non può essere un'operazione rigorosamente locale, ma dev'essere un processo globale implicante l'intero campo visivo o comunque gran parte di esso. In secondo luogo, il sistema visivo sembra supporre in effetti che l'intera immagine sia illuminata da una sola sorgente di luce. Questa particolarità potrebbe essere dovuta al fatto che l'evoluzione del cervello umano ha avuto luogo in un sistema planetario in cui esiste un unico sole.

Un'altra manifestazione di questa regola può essere vista in una forma complessa che fa pensare a un tubo bianco ad anse, illuminato da un lato (*c*). La forma appare quasi sempre convessa probabilmente a causa di sottili indizi, come l'occultamento di tratti del tubo, o a causa di una tendenza generale a vedere forme del genere come convesse. Fatto interessante, la profondità dei due dischi sovrapposti al tubo non è più ambigua: uno è chiaramente convesso e l'altro chiaramente concavo. Evidentemente, certi caratteri di un oggetto possono fornire al cervello informazioni sulla direzione dell'illuminazione, mentre la profondità di altre parti dell'oggetto viene di conseguenza adeguata alla sorgente di luce.

Il sistema visivo non solo suppone che ci sia una sorgente di luce singola, ma tende altresì a dare per scontato, abbastanza naturalmente, che la luce provenga dall'alto. Abbiamo dato un'illustrazione efficace di questo effetto per mezzo di una visualizzazione nella quale un gruppo di cerchi ombreggiati è semplicemente la versione capovolta di un altro gruppo (*si veda l'illustrazione a pagina 74 in basso*). I soggetti percepiscono invariabilmente il gruppo *a* come formato da sfere e il gruppo *b* come formato da cavità. Se si capovolge la pagina, si vedrà una sorprendente inversione della profondità: ora gli oggetti nel gruppo *b* appaiono convessi e quelli nel gruppo *a* concavi. (Il lettore può divertirsi a ritagliare l'illustrazione e applicarla sul piatto di un giradischi. Fino a quale velocità si può far girare il piatto del giradischi prima di smettere di vedere le inversioni?)

Il cervello supporrebbe dunque, come sembrano suggerire queste osservazioni, che il sole illumini i vari oggetti dall'alto. Ma in che modo il cervello riuscirebbe a distinguere fra «sopra» e «sotto»? Quel che importa è l'orientazione dell'oggetto in relazione alla retina, o piuttosto la sua orientazione rispetto al mondo esterno? Per valutare in modo adeguato questo punto, tentiamo il seguente esperimento: sdraiamoci su un divano e rovesciamo la testa oltre il bordo, in modo da guardare il mondo rovesciato. Ora chiediamo a un amico di mettersi dietro la nostra testa e di tenere dinanzi ai nostri occhi, nella sua corretta orientazione, l'illustrazione nella pagina seguente in basso. Gli oggetti nel gruppo *a* ci appariranno concavi e quelli nel gruppo *b* convessi; in altri termini si ha lo stesso effetto che si otterrebbe capovolgendo la pagina e osservandola tenendo la testa in posizione normale. Quel che conta è dunque l'orientazione dell'oggetto rispetto alla retina. La nostra conoscenza obiettiva di su e giù, alto e basso non incide sulla nostra percezione della profondità.

L'ombreggiatura genera di per sé solo una debole impressione di forma tridimensionale. Per trasmettere un'impressione convincente di profondità, la superficie ombreggiata dev'essere anche racchiusa da un contorno. In effetti, in molte delle nostre visualizzazioni la variazione di luminanza si avvicina solo in modo grossolanamente approssimato alla variazione continua (secondo il coseno) della vera ombreggiatura, e nondimeno già la semplice presenza di una linea di contorno circolare attorno alla regione in ombra può generare un'efficace illusione di sfericità. Ciò solleva un nuovo interrogativo: qual è il ruolo esatto del contorno nel determinare la percezione della forma considerando l'influenza dell'ombreggiatura?

Per rispondere a questa domanda, abbiamo disegnato una coppia di oggetti aventi uguale ombreggiatura, ma contorno diverso (*si veda l'illustrazione a pagina 75*). Le due immagini hanno lo stesso gradiente di luminanza: una foto-

L'interno di maschere cave, illuminato dall'alto, produce stranamente l'illusione di facce in rilievo illuminate dal basso. Nell'interpretazione delle immagini ombreggiate, il cervello tende a dare per scontato che la luce provenga dall'alto; in questo caso però respinge questo assunto per poter interpretare le immagini come normali oggetti convessi. Si osservi che i dischetti posti fra le due facce danno da parte loro l'impressione di essere illuminati dall'alto: il disco a destra appare convesso, e quello a sinistra concavo. Se si sovrappongono i dischi alla guancia (*a sinistra*), il loro rilievo diventa ambiguo. Se invece si «incorporano» i dischi nella guancia, sfumandone il contorno (*a destra*), essi vengono percepiti come se fossero illuminati dal basso, analogamente al resto della faccia.

cellula fatta scorrere su ciascuna immagine registrerebbe variazioni identiche nella distribuzione della luminanza. Eppure le due immagini sono notevolmente diverse fra loro. L'immagine superiore fa pensare a tre cilindri disposti fianco a fianco, mentre l'immagine inferiore trasmette l'impressione inconfondibile di un foglio di metallo ondulato. Le diverse percezioni sembrano dipendere in maniera pressoché totale dai contorni del bordo superiore e inferiore delle due superfici.

Da questi esempi concludiamo che, quando gli indizi forniti dall'ombreggiatura sono ambigui, le informazioni tratte dai bordi aiutano a risolvere l'ambiguità nell'intera immagine. Si riscontra inoltre un fatto piuttosto interessante: anche la posizione percepita della sorgente di luce si modifica per conformarsi alla superficie percepita. Nell'immagine superiore dell'ultimo esempio la luce sembra provenire da una sorgente in posizione perpendicolare rispetto alla pagina, mentre nell'immagine inferiore la luce proviene o dall'estrema sinistra o dall'estrema destra. È degno di nota che, modificando i confini di un oggetto, si possano produrre mutamenti così sorprendenti nella percezione.

Sfere o cavità? Dipende da dove si immagina collocata la sorgente di luce. Si può invertire la profondità (*a*) spostando mentalmente la sorgente da sinistra a destra. In una seconda disposizione di oggetti (*b*) ogni fila è di per sé ambigua poiché, ogni volta che se ne vede una come convessa, l'altra appare sempre concava. È quasi impossibile vederle entrambe simultaneamente come convesse o concave. La forma ad anse (*c*) fa pensare a un tubo bianco illuminato da destra. I due dischi su di esso sembrano conformarsi allo schema di illuminazione: il disco in alto appare convesso e quello in basso concavo. Quest'ultimo esperimento è stato realizzato dall'autore in collaborazione con Dorothy Kleffner e Steven J. Cobb.

Il cervello suppone che la luce provenga dall'alto. Perciò gli oggetti nel gruppo *a* appaiono convessi, mentre quelli nel gruppo *b* appaiono concavi. Se si capovolge la pagina, si rovescerà l'effetto di profondità degli oggetti. Rovesciando la testa e guardando la pagina tenuta diritta, si può dimostrare che ciò che conta è l'orientazione del disegno sulla retina.

L'esperimento seguente mostra che persino contorni illusori possono svolgere una funzione di questo genere. Un esempio tipico è costituito da quattro dischi in grigio scuro da cui sia stato tolto un pezzo (*si veda l'illustrazione a pagina 76 in alto*). Quando i dischi vengono disposti in modo appropriato, generano l'impressione di un grande disco chiaro al centro, che li copra parzialmente. In effetti, deboli linee sembrano connettere i margini concavi dei dischi grigi, anche se tali linee non esistono fisicamente.

Che cosa accade se sostituiamo lo sfondo di questa visualizzazione con uno in cui la luminanza, anziché essere uniforme, varia dall'alto verso il basso? Dapprima la nuova immagine così ottenuta offre l'impressione di essere piatta ma, se la si osserva per un po' di tempo, la regione all'interno del disco illusorio comincia a gonfiarsi verso l'osservatore e potrebbe addirittura staccarsi dallo sfondo per assumere l'aspetto di una sfera sospesa in aria. Abbastanza stranamente, un contorno illusorio sembra funzionare anche meglio di un contorno reale. La ragione non è del tutto chiara, ma il risultato suggerisce che il cervello consideri un fenomeno di occultamento come una prova più forte dell'esistenza di un oggetto rispetto alla presenza di un mero contorno. Dopo tutto, il contorno potrebbe rappresentare altrettanto bene una spira di sottile filo metallico o un'impalpabile bolla di sapone.

Quest'osservazione, al pari della precedente, dimostra l'esistenza di un'interazione diretta ed efficace fra i bordi, siano essi reali o illusori, e la derivazione della forma dall'ombreggiatura. Se il sistema visivo eseguisse misurazioni dettagliate della sola ombreggiatura per ricavarne l'orientazione delle superfici (come sembrano implicare alcuni modelli della visione elaborati nel campo dell'intelligenza artificiale), non si vedrebbe nell'immagine una sfera, giacché l'ombreggiatura non cambia affatto attraverso il bordo illusorio. Eppure il sistema visivo percepisce una sfera, perché l'ombreggiatura e il contorno illusorio concorrono a rinforzare una siffatta interpretazione.

Un altro modo in cui il sistema visivo giunge a delineare oggetti è attraverso variazioni nel potere riflettente della superficie, ossia nella proporzione di luce riflessa dalle superfici. Una fotocellula in movimento attraverso il confine di un oggetto registra di solito una variazione brusca di luminanza. Che cosa accadrebbe se il contorno fosse definito da un mutamento di colore, anziché da un mutamento di luminanza? Abbiamo preso una «sfera» ombreggiata nel modo già descritto e sostituito il fondo grigio uniforme con un fondo colorato in cui il gradiente di luminanza era uguale a quello della sfera. Il risultato è stato assai appariscente: l'illusione della profondità si è dissolta e la sfera è apparsa appiattita, anche se il suo contorno era ancora distintamente visibile a causa del contrasto di tonalità. Ne abbiamo concluso che il sistema sviluppato per ricavare la forma dall'ombreggiatura non può servirsi di bordi definiti da differenze di colore. Una ragione di ciò potrebbe risiedere nel fatto che i nostri primitivi progenitori primati, che assomigliavano ai tarsi, erano animali notturni e incapaci di distinguere i colori; nel loro mondo crepuscolare essi facevano affidamento, per percepire la profondità, sul solo contrasto di luminanza.

I margini influiscono sull'interpretazione delle superfici ombreggiate. Entrambe le immagini hanno la stessa variazione di ombreggiatura, ma l'immagine in alto dà l'idea di tre cilindri illuminati perpendicolarmente alla pagina, mentre quella in basso appare come un foglio metallico ondulato illuminato da una sorgente situata a destra (oppure a sinistra).

Tutti questi esempi implicano che il cervello ottenga informazioni sulla forma degli oggetti combinando contorni e indizi relativi all'ombreggiatura. Che cosa fa il cervello con le forme dopo averle determinate? Dal punto di vista percettivo è importante la capacità di isolare figure dallo sfondo. Persino in una scena ingombra di numerosi particolari il sistema visivo può decidere con facilità quali elementi dell'immagine siano legati fra loro così da formare oggetti. In una fotografia ad alto contrasto si può vedere un cane dalmata contro uno sfondo screziato (*si veda l'illustrazione nella pagina successiva in basso*). In modo analogo, si può staccare mentalmente un gruppo di linee che abbiano una particolare orientazione da un campo di linee con un'orientazione diversa. È d'altra parte impossibile separare mentalmente un gruppo di lettere rovesciate in modo speculare da un gruppo di lettere non rovesciate.

Le leggi del raggruppamento percettivo furono studiate sistematicamente per la prima volta, già parecchi anni fa, da Anne M. Treisman dell'Università della California a Berkeley, da Bela Julesz degli AT&T Bell Laboratories e da Jacob Beck dell'Università dell'Oregon. Questi ricercatori scoprirono un certo numero di importanti principi. In primo luogo, trovarono che una fase iniziale essenziale della percezione visiva implica l'estrazione di certi elementi fondamentali, che Julesz ha definito *textons*. Fra gli esempi si annoverano bordi orientati, colore e direzione di movimento. Una volta che il sistema visivo abbia estratto da un'immagine i caratteri elementari, i caratteri simili vengono raggruppati assieme a formare oggetti. In effetti, Beck suggerisce che solamente caratteri elementari, per definizione, possano essere raggruppati in questo modo. Presumibilmente, quindi, le lettere dell'alfabeto non costituiscono caratteri elementari, almeno per quanto concerne il sistema visivo.

Ma che cosa si può dire sugli oggetti tridimensionali? Gli esperimenti che sto per descrivere mostrano che persino forme definite esclusivamente per mezzo dell'ombreggiatura possono fungere da caratteri elementari per la percezione visiva. In una serie di cavità inframmezzate a forme convesse, per esempio, queste ultime possono essere mentalmente raggruppate assieme in modo da formare un piano di profondità autonomo chiaramente distinto dalle forme concave sullo sfondo (*si veda l'illustrazione a pagina 77*). Quando si osserva questa visualizzazione, si ha l'impressione che il sistema visivo passi per diverse fasi di elaborazione. Nella prima fase, il sistema esegue calcoli per definire le forme tridimensionali, impiegando per ottenere questo risultato vari secondi. Una volta che siano emerse le forme convesse, si ha la distinta impressione di essere in grado di controllarle indefinitamente per raggrupparle con forme simili nella visualizzazione. Infine, una volta che gli oggetti siano stati raggruppati, essi vengono chiaramente separati dalle forme non pertinenti, che fanno parte dello sfondo. L'estrazione e il raggruppamento dei *textons*, quindi, anche se di solito vengono descritti come un'operazione in un passo solo, possono implicare in realtà varie capacità percettive distinte che agiscono assieme per delineare una figura separandola dallo sfondo.

Ci siamo chiesti se il raggruppamento percettivo osservato in tale visualizzazione non potesse essere il risultato di qualche altro carattere, più elementare della forma tridimensionale. Per esempio, poiché le forme convesse differiscono da quelle concave nella polarità della loro luminanza da chiaro a scuro, si potrebbe supporre che il raggruppamento venga conseguito facendo riferimento alla polarità di luminanza. Per escludere questa possibilità, abbiamo creato un insieme di oggetti con le stesse polarità di luminanza di quelli della visualizzazione precedente, ma privi di ogni informazione sulla profondità. In questo insieme è virtualmente impossibile pervenire a un raggruppamento percettivo. Quand'anche si prendesse in considerazione ogni oggetto singolarmente, non si riuscirebbe a separarlo dal resto degli oggetti. Appare chiaro che i raggruppamenti osservati nella precedente visualizzazione devono fondarsi sulla forma tridimensionale piuttosto che su una polarità di luminanza.

Ho sottolineato che l'illusione della profondità è molto più efficace quando l'illuminazione proviene dall'alto che non quando sembra venire di lato. Similmente, l'illuminazione dall'alto aumenta di molto la capacità di raggruppare e separare immagini. Il lettore stesso può verificare quest'affermazione semplicemente ruotando di 90 gradi il gruppo *a* a pagina 77: l'impressione di profondità diminuirà e ci sarà una riduzione considerevole nella separazione percettiva. Questa osservazione conferisce un ulteriore sostegno all'idea che il raggrup-

Allineando opportunamente quattro dischi grigi da cui siano stati tolti alcuni pezzi si genera un cerchio illusorio (*a*). Uno sfondo ombreggiato crea una sfera illusoria (*b*). Un contorno reale non risulta altrettanto efficace (*c*). Un bordo circolare definito da un mutamento di luminosità crea la netta illusione di una sfera (*d*). Anche uno sfondo colorato definisce un bordo, ma poiché la luminosità dello sfondo corrisponde a quella del cerchio ombreggiato, l'illusione di profondità svanisce (*e*).

pamento percettivo debba fondarsi sulla forma tridimensionale. Questi raggruppamenti, inoltre, possono rappresentare di per sé forme di livello superiore, come un triangolo. Potrebbe essere interessante usare stimoli di questo genere per accertare se bambini piccoli o pazienti con danni cerebrali siano in grado di percepire forme sulla base dell'ombreggiatura; per esempio, un bambino piccolo risponderebbe a sfere disposte in modo da suggerire la forma di un viso?

Un'ulteriore notevole capacità della percezione visiva è quella di rilevare la simmetria. Questa capacità si estende anche a forme relativamente complicate, quali piante, facce e macchie di Rorschach. In che modo il sistema visivo scopre la simmetria? Paragona tutti i singoli elementi di un lato dell'immagine con quelli dell'altro lato per determinare se un oggetto sia simmetrico? Oppure raggruppa gli elementi in forme più significative per poi ricercare la simmetria in tali forme?

Nel tentativo di rispondere a queste domande, operammo un confronto fra due serie di cerchi ombreggiati (*si veda l'illustrazione a pagina 78 in alto*). I soggetti percepivano di solito il gruppo dell'illustrazione a sinistra come un insieme tridimensionale di sfere e cavità disposte simmetricamente rispetto a un asse orizzontale. Eppure, a un esame accurato (bidimensionale) si può rilevare che la metà inferiore della figura non è immagine speculare della metà superiore. In realtà è la disposizione a destra a essere davvero simmetrica. Questi risultati implicano che la percezione della simmetria si fondi sulla forma tridimensionale più che sulla semplice distribuzione di aree luminose e buie nell'immagine. Possiamo verificare quest'affermazione ruotando l'illustrazione di 90 gradi per eliminare la forte impressione di profondità. In questo modo saremo in grado di vedere che la disposizione di destra è più simmetrica della disposizione riportata a sinistra.

Le nostre osservazioni inducono a ritenere che nell'elaborazione del materiale visivo le informazioni fornite dalle ombre vengano estratte piuttosto presto. Di fatto, potrebbero esistere addirittura vie nervose deputate in modo specifico a questo compito. Recentemente, Terrence J. Sejnowski e Sidney R. Lekhy, della Johns Hopkins University, fondandosi su ricerche compiute con l'ausilio di un programma di simulazione al calcolatore, hanno prospettato la possibilità che cellule nervose del genere possano esistere effettivamente. Essi hanno iniziato a lavorare con una «rete neurale» formata da tre strati di neuroni e precisamente uno strato in ingresso, uno strato nascosto e uno strato in uscita. I neuroni degli strati in ingresso e in uscita simulavano il modello dei campi di neuroni recettivi circolari, disposti intorno al centro, che sono presenti nell'occhio del gatto. Un algoritmo di apprendimento adattava l'intensità dei segnali che passavano dai neuroni di uno strato a quelli dello strato successivo e dopo circa 40 000 tentativi la rete poteva correttamente associare forme ombreggiate ai rispettivi assi di curvatura tridimensionali.

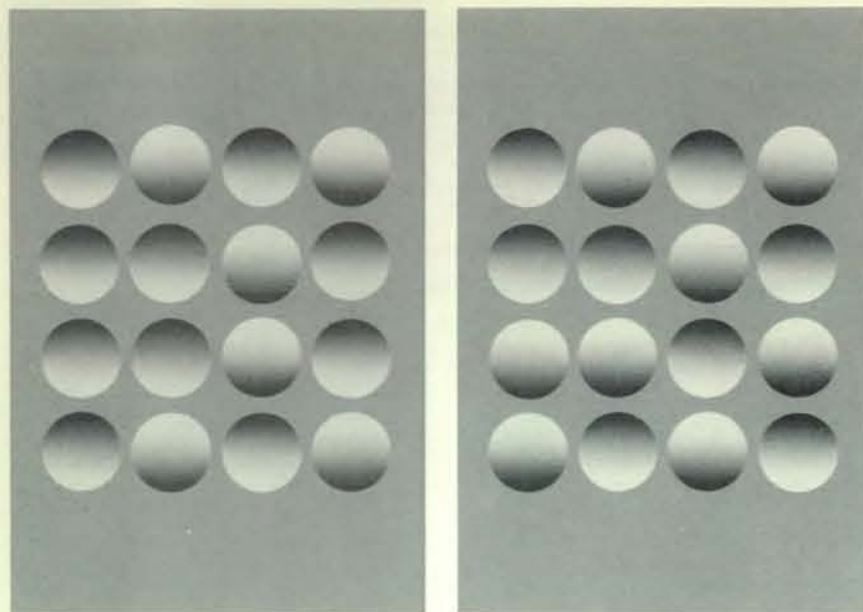
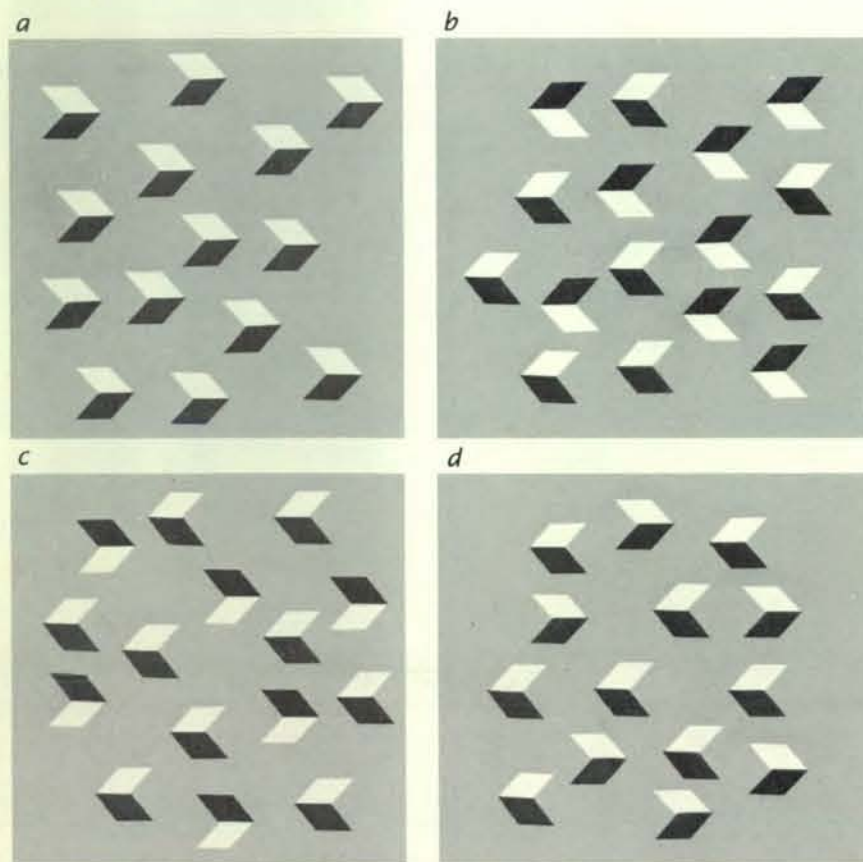
Quel che accadde poi fu una sorpresa: i ricercatori esaminarono le risposte delle cellule nello strato nascosto e trovarono che esse rispondevano a elementi lineari di varia lunghezza, larghezza e orientazione, rivelando una strana somiglianza con le cellule deputate al riconoscimento dei bordi che si trovano nella corteccia visiva di gatti e scimmie. Per quanto questa simulazione al calcolatore sia interessante, la sua pertinenza biologica non è ancora chiara, poiché i ricercatori esclusero deliberatamente profili e altri indizi che svolgono notoriamente una funzione cruciale nella visione umana. Rimane da vedere se la somiglianza fra le unità nascoste e i rivelatori corticali di bordi sia una semplice coincidenza o se le cellule rivelatrici di bordi servano veramente a estrarre forme tridimensionali dall'ombreggiatura.

a

b

c

Il raggruppamento percettivo di forme elementari permette di separare la forma di un cane dalmata da un fondo chiazzato in una fotografia ad alto contrasto (*a*). Le linee oblique possono essere raggruppate e considerate su un piano distinto da quello delle linee verticali (*b*). Le immagini speculari delle lettere non possono invece essere raggruppate visivamente e separate da quelle normali (*c*).

Il sistema visivo è in grado di distinguere forme convesse e concave e di raggrupparle assieme (*a*). In una disposizione con le stesse polarità di luminanza della precedente, ma priva di informazioni sulla profondità (*b*) è impossibile separare visivamente gli oggetti. In una disposizione illuminata di lato (*c*), il raggruppamento percettivo diventa più facile se si ruota l'immagine di 90 gradi. Gli oggetti convessi spiccano e formano un triangolo. L'ombreggiatura può definire forme complesse per un'ulteriore elaborazione visiva. Un'idea simile è stata avanzata da Alex Pentland del Massachusetts Institute of Technology.

La percezione della simmetria si verifica soltanto dopo che il cervello ha estratto informazioni sulla forma dall'ombreggiatura. Nell'immagine di sinistra sfere e cavità appaiono disposte simmetricamente attorno a un asse orizzontale. Se si considerano però le immagini in due dimensioni, è in realtà quella di destra a essere simmetrica.

I «galloni» nel gruppo *a* appaiono tutti o come cubi illusori o come «lapidi» che proiettano ombre, illuminati dallo stesso angolo. In *b* una regola di «puntamento» sembra soppiantare la regola della sorgente singola di luce e l'immagine viene vista come un insieme di cubi le cui facce abbiano riflettanza diversa. Il gruppo *c* viene sempre visto come mescolanza di cubi e lapidi perché quest'interpretazione soddisfa la regola della sorgente di luce singola. Nel gruppo *d*, ambiguo, è difficile unificare le figure in un'interpretazione coerente.

Finora ho preso in considerazione immagini stazionarie, ma che cosa è possibile dire degli oggetti in movimento? Si può scommettere con una ragionevole sicurezza sulla tesi che, in natura, pressoché tutto ciò che si muove è una preda o un predatore. Di conseguenza, il sistema visivo ha evidentemente sviluppato una grande varietà di meccanismi atti a scoprire il movimento. I dati empirici disponibili suggeriscono che la capacità di percepire il movimento sia mediata da gruppi di cellule cerebrali specializzate. Il meccanismo cerebrale che ci consente di percepire il moto può trarre vantaggio anche dall'informazione fornita dall'ombreggiatura? Per accertarlo decidemmo di sfruttare un'illusione ben nota, quella relativa al movimento apparente.

Possiamo produrre un esempio molto semplice di movimento apparente facendo lampeggiare in rapida alternanza due macchie di luce spazialmente separate. Anziché vedere due luci che si accendono e che si spengono, di solito si vede una singola luce che salta avanti e indietro. Per determinare il ruolo giocato dagli indizi forniti dall'ombreggiatura nella percezione umana del movimento, approntammo una visualizzazione che passava rapidamente da un'immagine che mostrava un oggetto ombreggiato convesso al di sopra di uno concavo a una seconda immagine in cui i due oggetti erano scambiati. Undici soggetti non addestrati riferirono di aver visto una sfera che saltava su e giù fra due buchi nello sfondo.

Questo risultato suggerisce che il cervello prima di poter percepire il movimento apparente debba valutare la forma tridimensionale. Di fatto i soggetti impiegano spesso decine di secondi per sviluppare un'impressione di profondità e durante questo tempo non vedono alcun movimento apparente. Pare pertanto improbabile che il movimento apparente possa fondarsi su un qualche altro elemento più primitivo dell'immagine. Per ottenere un'illustrazione più diretta di questa tesi, ruotammo l'intera immagine di 90 gradi. In questo modo risultò considerevolmente ridotta l'impressione di profondità e si giunse a una perdita quasi completa dell'effetto di movimento apparente.

Sembra, quindi, che il nostro sistema visivo estragga un oggetto tridimensionale da indizi fondati sull'ombreggiatura e che percepisca il movimento sulla base dell'immagine tridimensionale, piuttosto che grazie all'uso diretto dell'immagine «primitiva» bidimensionale. Certe cellule nella corteccia visiva delle scimmie rispondono al movimento apparente di stimoli semplici, come le luci lampeggianti descritte sopra. Potrebbe essere interessante vedere se queste cellule siano o meno in grado di rispondere a un movimento fondato su oggetti la cui forma venisse percepita sulla base dell'ombreggiatura.

È chiaro che la percezione visiva si fonda su una vasta gamma di processi biologici per pervenire a una rappresentazione tridimensionale del mondo. Per creare questa rappresentazione, il sistema visivo sembra presupporre una varietà di ipotesi semplificatrici, come la regola in base alla quale vi sarebbe solo una sorgente di luce. Che cosa accade quando l'apparato visivo cerca di costruire una scena coerente mettendo assieme molti frammenti disparati? Recentemente Patrick Cavanagh, Diane Rogers-Ramachandran e io abbiamo effettuato uno studio per rispondere a questa domanda.

Abbiamo creato disposizioni semplici e casuali di forme simili a «galloni» da sergente, ciascuna delle quali può essere vista come una coppia di facce adiacenti di un cubo (*si veda l'illustrazione nella pagina a fronte*). La disposizione *a* può essere percepita come una serie di cubi paralleli orientati tutti nella stessa direzione e illuminati da una singola sorgente di luce; i parallelogrammi neri sono visti come la faccia in ombra del cubo. Altrettanto spesso, però, la disposizione è percepita come un insieme di «lapidi tombali» bianche che proiettano ombre nere. Spostando mentalmente la direzione della sorgente di luce si può passare dalla percezione di cubi a quella di lapidi. Si osservi che quando, in uno di questi gruppi di figure, se ne vede una come un cubo, anche tutte le altre appaiono come cubi. È in effetti impossibile percepire simultaneamente alcune figure come cubi e altre come lapidi, poiché una tale percezione violerebbe la regola della sorgente di luce singola. Fatto interessante, quando le forme sono percepite come cubi, c'è una tendenza a integrare la faccia mancante: ossia a percepire superfici illusorie. Le superfici illusorie svaniscono quando le forme sono percepite come lapidi.

Successivamente abbiamo capovolto o rovesciato metà circa dei galloni in combinazioni diverse. Queste nuove immagini illustrano la sottile interazione di costruzioni e di regole di organizzazione che si verifica quando il cervello tenta di creare forme significative da frammenti isolati. Nella disposizione *b*, per esempio, tutte le forme appaiono di solito come cubi paralleli, anche se ciò sarebbe incompatibile con la condizione della sorgente di luce singola. A quanto pare, quando la regola della sorgente di luce singola non può essere soddisfatta, viene sostituita da una regola di «puntamento» (ovvero, da una regola la quale dice che forme con orientazioni simili sono in realtà superfici parallele). Al fine di evitare conflitti, il cervello suppone semplicemente che i cubi abbiano facce di colore diverso.

Nella disposizione *c* si osserva un miscuglio di lapidi e di cubi, poiché questo fatto permette al sistema di soddisfare la regola della sorgente di luce singola. È in realtà impossibile vedere l'insieme come formato esclusivamente da cubi o da lapidi, poiché una tale percezione non sarebbe compatibile né con la regola del puntamento, né con la regola della sorgente di luce. Ci si troverà inoltre costantemente impegnati a unificare tutte le forme dell'immagine in una singola superficie coerente, in modo da suggerire una superficie metallica incisa con «gradini» situati a caso scolpiti in essa. Mentre una formica che camminasse sulla figura percepirebbe solo variazioni caotiche di luminosità, l'occhio umano passa in rassegna l'intera immagine e unisce le superfici parallele in un tessuto compatto per creare ordine e unità spaziali.

Nella disposizione *d* le figure non sono né parallele, né in grado di soddisfare la costrizione della singola sorgente di luce. Qui c'è una tendenza a vedere nell'immagine un raggruppamento casuale di galloni piani tendenti in direzioni opposte. Benché una singola figura dell'insieme possa essere vista talvolta come un cubo o come una lapide, è difficile unificare tutte le forme in un'interpretazione tridimensionale coerente.

Nel mondo reale l'immaginazione visiva - ossia una conoscenza ad alto livello circa ciò che si sta vedendo - esercita un'influenza profonda sulla percezione della forma in base all'ombreggiatura. Di fatto l'interazione fra immaginazione visiva e percezione è uno fra gli argomenti più elusivi ed enigmatici in psicologia. Per chiarire questo punto creammo una disposizione di cerchi ombreggiati che, a un'ispezione casuale, sembrano formare serie alternate di sfere e cavità (*si veda l'illustrazione in questa pagina*). La disposizione è però suscettibile di un'interpretazione radicalmente diversa: essa può essere vista come un foglio in cui siano stati praticati 16 fori, dietro cui si intravvedono due bande scure dai bordi sfocati. Questo mutamento di percezione fa perdere completamente ai cerchi ombreggiati la loro forma sferica.

La tendenza a vedere bande anziché sfere e cavità può essere accentuata da indizi stereoscopici. Nell'illustrazione *b* in questa pagina si possono vedere delle sfere, ma se dovessimo «fonderle» in una visione binoculare per mezzo di un visore stereoscopico, vedremmo un'immagine con una finestra circolare stagliarsi chiaramente da uno sfondo in ombra. Diventerebbe in effetti virtualmente impossibile vedere la figura circolare come una sfera anziché come un buco. Questo implica che l'estrazione della forma dall'ombreggiatura è un processo che risente in misura molto marcata dell'elaborazione stereoscopica.

L'interpretazione della forma sulla base dell'ombreggiatura interagisce fortemente anche con la conoscenza degli oggetti che il sistema visivo dà per acquisita, come dimostra nel modo più chiaro l'illustrazione di apertura di questo articolo. Nelle due immagini, l'interno cavo della maschera è illuminato dall'alto; ci si dovrebbe perciò attendere di vedere due figure cave. Il sistema visivo tende però fortemente a respingere la possibilità che le forme siano cave e interpreta perciò queste immagini come normali facce in rilievo. Esso «aggira» pertanto il consueto assunto in base al quale l'illuminazione proverrebbe dall'alto per poter interpretare le forme come facce normali.

a

b

L'immaginazione visiva, o un'informazione di livello superiore sugli oggetti, influisce profondamente sulla percezione della forma in base all'ombreggiatura. Le file di sfere e di cavità (*a*) possono essere interpretate anche come due bande scure dai bordi sfumati viste attraverso 16 buchi in un foglio opaco. In tal caso non si possono più vedere le forme sferiche. Ognuna delle due immagini in *b* rappresenta una sfera, ma quando le due immagini vengono «fuse» assieme in un visore stereoscopico, le sfere svaniscono e si vede una finestra circolare aperta in un foglio rettangolare sospeso davanti a un piano ombreggiato.

Si osservino ora i due piccoli dischi ombreggiati raffigurati fra le due facce. Benché si supponga che la luce che illumina le due facce provenga dal basso, il disco di destra viene visto normalmente come convesso e quello di sinistra come concavo, ovvero vengono percepiti entrambi come illuminati dall'alto. Forse il cervello tratta questi oggetti come del tutto distinti dalle facce e perciò, nell'interpretarne l'ombreggiatura, aderisce alla regola più «primitiva», supponendo cioè che essi siano illuminati dall'alto. Se però si applicano i dischi a una guancia di una delle due facce, il loro rilievo diviene ambiguo: il disco di destra può apparire concavo e quello di sinistra convesso. Infine, se si «incorporano» i dischi sulla faccia eliminandone il contorno, questi vengono percepiti invariabilmente come illuminati dal basso, analogamente al resto della faccia. Di conseguenza il disco a destra appare come una cavità, mentre quello a sinistra fa pensare a un bernoccolo.

La nostra ricerca ha consentito di individuare una varietà di regole che vengono applicate in una fase molto precoce dell'elaborazione visiva della forma in base all'ombreggiatura. Abbiamo osservato che è possibile ricostruire il flusso di informazione dalle primissime fasi della percezione della forma sino alla fase finale, dove l'informazione interagisce con una conoscenza ad alto livello delle sorgenti di luce e della natura di complessi oggetti tridimensionali. Gli eventi cerebrali che mediano l'elaborazione dell'informazione visiva negli esseri umani sono ancora misteriosi, ma la psicologia può aiutarci a chiarire quali possano essere questi eventi e in che modo siano organizzati nel cervello. Anche nuovi modelli sviluppati con l'aiuto del calcolatore possono aiutarci a inquadrare la ricerca. Questi sviluppi stanno proiettando le indagini sulla percezione visiva in un nuovo ambito, nel quale un giorno si potranno forse spiegare i meccanismi cellulari cerebrali che ci consentono di percepire visivamente il mondo in tre dimensioni.

BIBLIOGRAFIA

YONAS ALBERT, KUSKOWSKI MICHAEL e STERNFELS SUSAN, *The Role of Frames of Reference in the Development of Responsiveness to Shading* in «Child Development», 50, n. 2, giugno 1979.

TODD JAMES T. e MINGOLLA ENNIO, *Perception of Surface Curvature and Direction of Illumination from Patterns of Shading* in «Journal of Experimental Psychology: Human Perception and Performance», 9, n. 4, agosto 1983.

RAMACHANDRAN V. S., *Perception of Shape from Shading* in «Nature», 331, n. 6152, 14 gennaio 1988.

La visualizzazione di sorgenti che emettono raggi X

Una variante dello stenoscopio dotata di molti forellini disposti secondo schemi particolari consente la realizzazione di maschere di codifica per visualizzare sorgenti di raggi X di alta energia come i plasmi e i buchi neri

di Gerald K. Skinner

Come si può fotografare un oggetto lontano che emette solo raggi X? Telescopi e apparecchiature che usano comuni obiettivi e specchi non servono a nulla. Infatti i fotoni dei raggi X molli (di bassa energia), a differenza di quelli della luce visibile, hanno energia sufficiente per rimuovere gli elettroni dagli atomi del materiale con cui sono fatte le lenti e quindi vengono fermati dalle lenti stesse. I raggi X duri (di alta energia) possono attraversare un obiettivo ma, dato che virtualmente non subiscono alcuna deflessione, non possono venire messi a fuoco. I raggi X molli possono invece essere focalizzati, ma solo se vengono fatti incidere successivamente su più superfici riflettenti, con angoli di incidenza molto piccoli. La riflessione a incidenza radente diventa però sempre meno efficiente all'aumentare dell'energia dei fotoni, con la conseguente diminuzione dell'angolo di riflessione. Un telescopio per raggi X di alta energia richiederebbe quindi superfici riflettenti molto grandi e la loro realizzazione, con la necessaria precisione e con il grado di finitura richiesto, risulterebbe costosa in maniera proibitiva. Questa tecnica di visualizzazione diviene pertanto inutilizzabile con raggi X di energia superiore a 10 000 elettronvolt, corrispondente a una lunghezza d'onda inferiore a circa 0,12 nanometri (miliardesimi di metro).

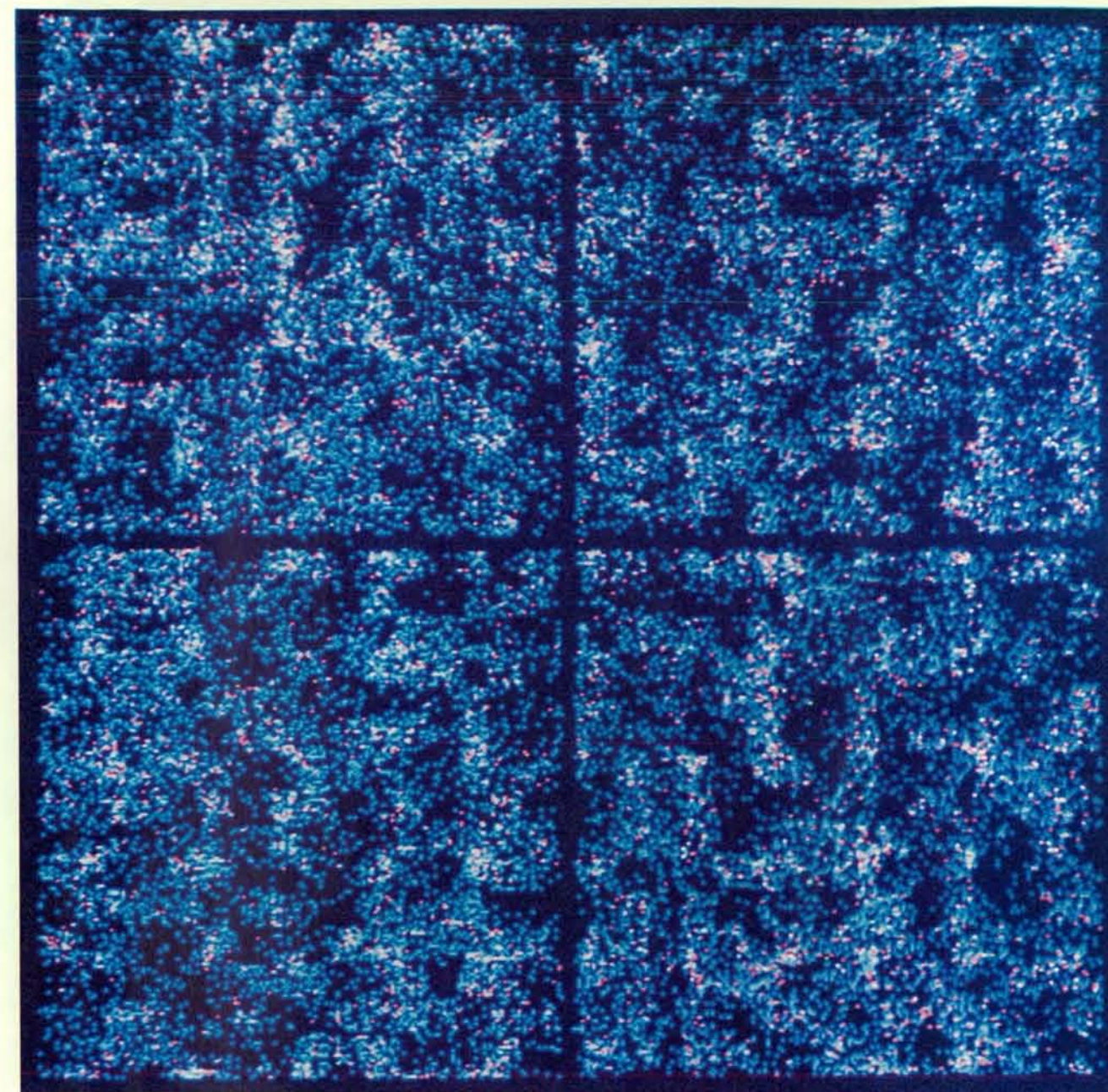
Com'è possibile allora ottenere immagini di oggetti che emettano raggi X di elevata energia? In effetti la cosa è possibile usando una tecnica simile a quella impiegata per fare le comuni radiografie cliniche. Queste vengono eseguite ponendo una pellicola con emulsione sensibile ai raggi X immediatamente dietro alla parte del corpo che viene irradiata con una debole sorgente puntiforme di raggi X. La pellicola registra di fatto l'ombra proiettata dalle parti del corpo che assorbono i raggi X. Con un termine ormai in disuso tali radiografie vengono dette infatti anche skiadografie, dal greco *skia*, che significa ombra.

Ma supponiamo ora che i raggi X non provengano da una sorgente puntiforme come un tubo radiografico, bensì da un oggetto esteso di forma sconosciuta, come un grande nube di plasma intergalattico caldo. Dato che una sorgente estesa può venire considerata come l'insieme di moltissime sorgenti puntiformi, l'ombra proiettata da un oggetto interposto è mal definita, perché ogni punto della sorgente proietta sulla pellicola un'ombra leggermente diversa. Tuttavia, se si conosce con esattezza la forma dell'oggetto interposto, si può facilmente prevedere quale sarebbe la forma dell'ombra proiettata qualora esso venisse «illuminato» con una sorgente puntiforme. Confrontando le ombre prodotte da tutte le possibili combinazioni di sorgenti puntiformi con quella effettivamente registrata dovrebbe allora risultare possibile determinare la forma della sorgente estesa.

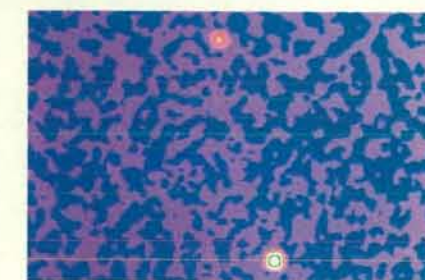
Su questo principio si basa in effetti la visualizzazione di sorgenti X mediante maschere di codifica dell'immagine. L'oggetto di forma nota interposto fra la sorgente e il rivelatore è una maschera ritagliata secondo un codice noto, in pratica una lastra opaca ai raggi X con fori opportunamente disposti. La chiave di questo tipo di visualizzazione sta nella scelta di una disposizione dei fori che consenta la ricostruzione dell'immagine della sorgente di raggi X a partire dalla forma dell'ombra della maschera.

La maschera più semplice possibile, quella con un solo piccolo foro, è un caso particolare che aiuta a comprendere come si procede in un caso più generale. In effetti tale sistema è uno stenoscopio (*si veda l'illustrazione a pagina 86*). L'ombra a valle della maschera sarebbe in questo caso una immagine invertita e non proprio nitidissima della sorgente di radiazione. Un'apparecchiatura di questo tipo potrebbe funzionare con raggi X duri, ma è di scarsa utilità pratica. Un'immagine nitida può essere ottenuta solo usando un foro molto piccolo, ma in questo caso la quantità di radiazione che viene registrata è pure piccola. Una maschera con un foro di un millimetro di diametro orbitante intorno alla Terra e illuminata da una tipica sorgente cosmica di raggi X lascerebbe passare circa un fotone X (un quanto di radiazione) all'ora. Dato che il livello del rumore di fondo (prodotto dai raggi cosmici) registrato da un rivelatore di raggi X è ben superiore a un quanto all'ora, una maschera con un solo foro non costituisce una soluzione adatta alla ripresa di immagini delle sorgenti cosmiche di raggi X.

Un rimedio ovvio consiste nel fare più fori nella maschera, dato che il numero di raggi X rilevati a valle di essa è proporzionale al numero dei fori. Sfortunatamente ogni foro produce una propria immagine invertita e poco nitida, cosicché il rivelatore «vede» un gran numero di immagini sovrapposte. Come è possibile ottenere da quest'ombra una sola immagine nitida?

A questo fine occorre innanzitutto dividere il campo inquadrato in un gran numero di elementi o pixel. A ognuno di questi si può attribuire un livello di luminosità proporzionale al flusso di raggi X che provengono da quell'elemento. Ogni pixel produce sostanzialmente la medesima ombra della maschera, ma le ombre prodotte da ciascuno di essi sono traslate rispetto a quelle prodotte dai rimanenti pixel. Con semplici considerazioni geometriche è possibile determinare le posizioni relative delle ombre prodotte da ciascun pixel che emetta raggi X.

Dato che una tipica sorgente di raggi X può venire adeguatamente rappresentata solo con un gran numero di pixel, l'ombra che viene registrata dal rivelatore è quindi la sovrapposizione di molte versioni traslate della medesima ombra. Nondimeno è possibile correlare l'ombra complessa registrata a valle della maschera e quella che sarebbe prodotta da ciascun pixel se fosse la sola sorgente di raggi X. Una forte somiglianza fra le due ombre (o, per essere più precisi, un elevato coefficiente di correlazione) indica che il pixel considerato è una intensa sorgente di raggi X. Una bassa correlazione dimostra invece che quel pixel contribuisce poco o nulla al flusso di raggi X.

Considerando in questo modo, uno dopo l'altro, tutti i pixel si può dunque costruire un'immagine della sorgente. Ma, dato che si ha a che fare con decine di migliaia di fotoni registrati e con un numero di pixel circa dello stesso ordine di grandezza, i calcoli richiesti per ottenere un'immagine non sono semplici. Fortunatamente le tecniche numeriche che utilizzano le cosiddette trasformate di Fourier e di Hadamard abbreviano in modo significativo il tempo necessario per i calcoli e consentono la produzione di questo tipo di immagini.

Naturalmente una maschera potrebbe produrre ombre molto simili anche quando fosse illuminata da raggi X provenienti da due pixel diversi. In tal caso nell'immagine ricostruita entrambi apparirebbero luminosi, anche se uno solo di essi potrebbe aver effettivamente emesso raggi X. Occorre pertanto scegliere con cura la disposizione dei fori

I dati raccolti da un telescopio a maschera di codifica dell'immagine possono presentarsi in forma «divisionista». Ogni punto dell'immagine in alto indica dove un fotone X è caduto sul rivelatore. (La croce è dovuta a una serie di montanti sopra il rivelatore). Elaborando i dati al calcolatore si ottiene l'immagine (*in basso*) di due sorgenti celesti di raggi X: GX3 + 1 e GX5 − 1.

nella maschera, in modo che l'ombra prodotta quando la maschera viene illuminata da un pixel risulti il più possibile diversa da quella prodotta dalla radiazione di ogni altro pixel. In altre parole l'ombra proiettata da una maschera deve avere una bassa correlazione con qualsiasi sua altra versione traslata.

Una disposizione regolare dei fori costituirebbe di conseguenza una scelta poco felice. La simmetria della maschera rispetto alla traslazione potrebbe produrre ombre virtualmente indistinguibili dalle loro versioni traslate, introducendo quindi una notevole ambiguità nella determinazione della provenienza dei raggi X rilevati. Un'alternativa molto migliore sarebbe costituita da una disposizione del tutto casuale dei fori sulla superficie della maschera e, in effetti, simili maschere sono state usate. Eppure una distribuzione casuale dei fori non rappresenta ancora la scelta migliore.

Il problema della scelta della conformazione di una maschera assomiglia a problemi familiari in altri campi della scienza e della tecnica. Per esempio i codici numerici usati per la trasmissione di informazioni lungo linee telefoniche rumorose devono essere quanto più possibile diversi l'uno dall'altro, in modo che, se vengono alterati alcuni bit di un messaggio, non si riceva un codice del tutto errato. Anche le mire usate per verificare l'allineamento dei sistemi ottici devono essere di foggia tale da risultare facilmente riconoscibili da una loro versione traslata.

Le migliori soluzioni per problemi di questo tipo sono date da enti matematici detti «insiemi alle differenze cicliche», la cui natura si spiega meglio con un esempio. Disponiamo lungo un cerchio i numeri interi da 1 a 15 (*si veda l'illustrazione a pagina 88*) in modo che contando 15 passi unitari a partire da 1 si ritorni a 1. Fra i numeri così disposti, 1, 2, 3, 5, 6, 9 e 11 hanno fra di loro una particolare relazione. In tre casi i numeri di questo insieme sono adiacenti (1 e 2, 2 e 3, 5 e 6) e in altri tre casi sono separati da un passo unitario (1 e 3, 3 e 5, 9 e 11). In effetti, a un esame più approfondito, si vede che vi sono sempre tre coppie di numeri separate da un passo grande quanto si vuole, purché non multiplo di 15. Quindi, se si sostituiscono i numeri elencati con dei fori e si ruota il cerchio di un numero qualsiasi di passi interi, in generale solo tre fori assumeranno una posizione coincidente con la posizione iniziale di altri fori. Non è una gran sovrapposizione, considerando che vi sono sette fori. Le eccezioni si hanno quando si ruota il cerchio di passi multipli di 15, nel qual caso si otterrà sempre una sovrapposizione totale dei fori.

Queste proprietà fanno sì che i numeri 1, 2, 3, 5, 6, 9 e 11 costituiscano un insieme alle differenze cicliche «in base 15». Vi sono altri insiemi di questo tipo con basi differenti, che forniscono numeri diversi di fori coincidenti per rotazione, ma tutti possiedono la medesima proprietà: ruotando, o spostando, ciclicamente i membri dell'insieme si ha sempre una cattiva sovrapposizione, a meno che il passo non sia un multiplo della base.

Per produrre una maschera a partire da un insieme in base n si comincia con una griglia regolare composta da linee verticali e orizzontali ugualmente spaziate. Si numerano poi da 1 a n i quadrati formati dall'intersezione delle righe. Tuttavia per essere certi che le proprietà appena viste vengano trasferite in un insieme bidimensionale occorre numerare la griglia diagonalmente. Si praticano poi fori in corrispondenza dei numeri che appartengono all'insieme.

In pratica, la griglia di fori è prodotta da un calcolatore, che la ripete quattro o più volte e poi la trasferisce su una pellicola fotografica. Una maschera con una complessa codifica può essere facilmente costruita proiettando la pellicola su una lastra metallica rivestita da una sostanza sensibile alla luce. Il metallo in corrispondenza delle zone nelle quali la sostanza fotosensibile è stata distrutta dall'esposizione alla luce (ossia le zone in cui dovranno trovarsi i fori) viene poi sciolto con un acido. Per essere sicuri che il metallo rimasto sia veramente opaco ai raggi X, a volte è necessario costruire la piastra con un metallo raro, come il tantalio. In alternativa, si può ricorrere alla deposizione elettrolitica di un buon assorbitore di raggi X, come l'oro. Questo metodo è stato usato per la produzione di maschere con più di 32 000 fori.

SORGENTE DI RAGGI X

MASCHERA

RIVELATORE

Il più semplice sistema di visualizzazione con maschera di codifica (*in alto*) è costituito da un rivelatore posto dietro a una maschera, ossia una lastra opaca ai raggi X munita di un solo forellino. Il sistema si comporta come uno stenoscopio. Solo i raggi X che convergono nel foro giungono al rivelatore, dove producono un'immagine invertita e non del tutto nitida della sorgente. In pratica, dato che il numero di quanti X che attraversano il foro è piccolo, il sistema è di scarsa utilità per ottenere immagini di sorgenti di raggi X deboli o lontane. Una maschera con più fori lascia passare più quanti X (*in basso*), ma ogni foro produce un'immagine leggermente traslata della sorgente; è quindi necessario ricorrere all'elaborazione elettronica per ricostruire l'immagine della sorgente codificata dalla maschera.

Ma come si ottiene l'insieme alle differenze cicliche in base n che serve per produrre la maschera? Il fatto è che tali insiemi vanno cercati fra tutti i numeri interi compresi tra 0 e n, per quanto vi siano alcune regole generali che aiutano a trovarli. Sotto questo punto di vista tali insiemi rassomigliano ai numeri primi e, in effetti, alcuni metodi per trovarli richiedono innanzitutto la determinazione di una coppia di numeri primi. Sono state identificate varie famiglie di insiemi alle differenze cicliche, che forniscono maschere estremamente diverse l'una dall'altra (*si veda l'illustrazione a pagina 90*). Alcune rassomigliano a schemi di parole incrociate, altre a opere di op-art. Alcune hanno fori molto radi, in altre questi occupano circa metà delle possibili posizioni.

Per la registrazione delle ombre proiettate dalle maschere si possono usare pellicole simili a quelle usate per le radiografie cliniche, ma esse presentano diversi svantaggi. Innanzitutto non sono sensibili a flussi molto deboli di raggi X e non possono distinguere raggi X di diverse energie né, tantomeno, raggi X e raggi cosmici di fondo. Inoltre le emulsioni a base di sali d'argento raggiungono presto la saturazione: dopo che un grano dell'emulsione è stato colpito da un quanto di radiazione X non risponde più ad altri quanti che eventualmente lo raggiungano.

Per questi motivi l'ombra della maschera viene spesso registrata con rivelatori denominati contatori proporzionali a riempimento gassoso, che possono venire impiegati anche per registrare le immagini prodotte dai telescopi X a incidenza radente. Quando un fotone X incide su un contatore proporzionale, questo emette un impulso elettrico la cui ampiezza dipende dall'energia del fotone. I contatori proporzionali sono molto più sensibili dei materiali usati in radiografia e la loro capacità di discriminare le energie dei raggi X consente di studiare lo spettro di emissione della sorgente. Inoltre è possibile misurare le piccole differenze fra i tempi di arrivo dell'impulso allo due estremità del rivelatore, cosa che consente di determinare con una precisione superiore a un millimetro in che punto il fotone X ha colpito il rivelatore.

I componenti fondamentali di un sistema di visualizzazione di sorgenti di raggi X sono quindi una maschera di codifica, un contatore proporzionale sensibile alla posizione di incidenza e un sistema di elaborazione dei dati. La maschera viene posta davanti al rivelatore a una distanza appropriata e l'insieme viene puntato verso la sorgente da riprendere. Si registra la distribuzione dei raggi X che incidono sul rivelatore e, se la maschera è stata scelta opportunamente, la sua ombra contiene abbastanza informazioni per ricostruire al calcolatore l'immagine della sorgente.

Un tipico sistema di visualizzazione con maschera di codifica (*in alto*) usa una maschera con una disposizione dei fori dettata da un «insieme alle differenze cicliche» (*si veda l'illustrazione a pagina 88*). Illuminando tale maschera con una sorgente di raggi X, l'ombra proiettata contiene, codificata in maniera univoca, la forma della sorgente. I raggi X emessi della sorgente producono infatti sul rivelatore ombre che possono essere distinte l'una dall'altra, sebbene siano solo versioni traslate della medesima ombra fondamentale. Al contrario le ombre proiettate da una maschera con fori disposti regolarmente (*in basso*) sono virtualmente identiche, qualunque sia il punto dal quale viene illuminata la maschera. Nell'illustrazione sono mostrate solo due sorgenti puntiformi di raggi X, ma una sorgente estesa può considerarsi formata da numerosissime sorgenti, ciascuna con la propria ombra.

È interessante notare che l'insieme maschera-rivelatore funziona un po' come un obiettivo zoom per fotografia, dato che variando la distanza fra i due si ottengono diverse combinazioni di potere risolvente e di angolo di campo. Il medesimo sistema può servire inoltre sia come telescopio per visualizzare sorgenti lontane di raggi X, sia come «apparecchio fotografico» per riprendere oggetti vicini. In entrambi i casi la registrazione viene compiuta nello stesso modo. La sola differenza consiste in piccole variazioni dei fattori numerici usati per l'elaborazione. In pratica la messa a fuoco di un sistema di visualizzazione con maschera di codifica avviene dopo aver ripreso l'immagine.

Una delle applicazioni di questi sistemi di visualizzazione si ha nel campo della fusione nucleare, dove si studiano le proprietà dei plasmi a temperature elevatissime. Quando il plasma raggiunge una temperatura prossima a quella ne-

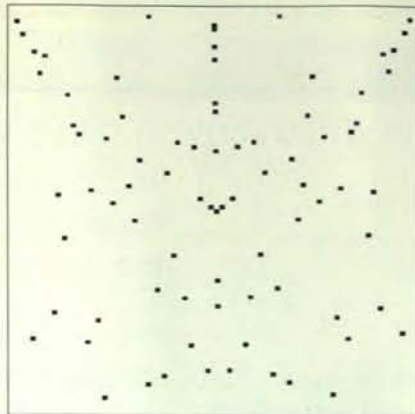
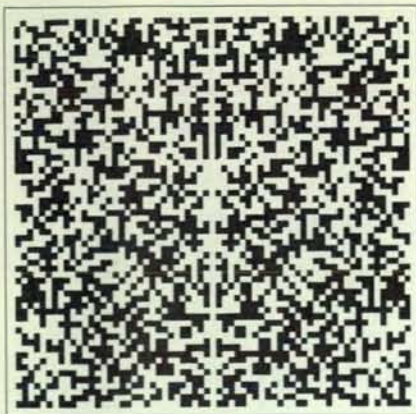
L'insieme alle differenze cicliche in base 15 è formato dai numeri 1, 2, 3, 5, 6, 9 e 11, evidenziati in colore nel cerchio in alto, che contiene i numeri da 1 a 15. Se il cerchio viene fatto ruotare di due passi unitari, solo in tre casi i numeri dell'insieme vengono a trovarsi in posizioni precedentemente occupate da altri numeri del medesimo insieme (*archi interni*). Lo stesso vale per una rotazione di tre passi (*archi centrali*) o di sei (*archi esterni*). In generale, vi saranno solo tre sovrapposizioni per qualsiasi rotazione, a meno che il passo non sia multiplo di 15, nel qual caso si avrà una sovrapposizione completa. Questa proprietà rende tali insiemi molto utili per la costruzione di maschere per la codifica di immagini. Infatti ruotando, o traslando, i membri dell'insieme si ha una corrispondenza imperfetta con le posizioni prima occupate, cosa che rende facilmente distinguibili le diverse ombre proiettate da maschere con fori corrispondenti alle posizioni dei numeri dell'insieme. Per realizzare una maschera di codifica da un insieme alle differenze cicliche in base *n*, vengono tracciate linee verticali e orizzontali che formino una scacchiera con *n* caselle (*in basso*). Tali caselle vengono numerate diagonalmente e infine vengono praticati i fori (relativi ai punti dell'insieme) mediante un procedimento fotolitografico. La disposizione di base viene ripetuta quattro o più volte per formare una maschera completa.

cessaria per la fusione, l'energia tende a venire irradiata sotto forma di raggi X. Ottenendo un'immagine della radiazione emessa, è pertanto possibile osservare le dimensioni e la forma del plasma e determinare le sue variazioni di temperatura. Altre applicazioni si hanno in medicina, dove occorre spesso visualizzare l'assorbimento e la distribuzione nell'organismo di farmaci marcati con radionuclidi (che emettono raggi X e gamma).

Ma è in astronomia che la nuova tecnica di visualizzazione di raggi X con maschere di codifica ha avuto le maggiori applicazioni. Strano a dirsi, gli astronomi sono particolarmente interessati ai raggi X cosmici proprio perché la maggior parte delle stelle e degli altri oggetti dell'universo non emettono raggi X in maniera cospicua. La scoperta che alcuni oggetti astronomici emettono grandi quantità di raggi X è stata una sorpresa relativamente recente perché l'astronomia dei raggi X ha dovuto attendere la diffusione di palloni d'alta quota, razzi e satelliti che potessero portare i sistemi di visualizzazione al di sopra dell'atmosfera terrestre, che assorbe i raggi X.

Quando si scopre che da una certa zona del cielo giungono raggi X, se ne deduce che laggiù sta succedendo qualcosa di fuori del comune. Significa infatti che vi sono plasmi a elevatissime temperature (almeno qualche milione di gradi) o che particelle vengono accelerate a energie molto maggiori di quanto avvenga normalmente nei pressi delle stelle o di altri comuni corpi celesti. La lista delle sorgenti di raggi X attualmente note comprende molti degli oggetti più strani dell'universo: supernove, stelle di neutroni, buchi neri, plasmi intergalattici ultracaldi e i nuclei dei quasar.

La maggior parte delle immagini di sorgenti di raggi X ottenute fino a oggi sono state realizzate con telescopi dotati di ottiche a incidenza radente (si veda l'articolo *L'osservatorio Einstein* di Riccardo Giacconi in «Le Scienze» n. 140, aprile 1980) e quindi sono per lo più relative a sorgenti di raggi X molli. Prima dello sviluppo dei sistemi di visualizzazione a maschera di codifica, era possibile studiare solo lo spettro e le variazioni di intensità delle sorgenti di raggi X duri. Oggi gli astronomi possono studiare alcuni fra i processi di più alta energia a noi noti ottenendo fotografie a raggi X del cielo con telescopi a maschera di codifica.

La possibilità di far funzionare nello spazio uno strumento a maschera di codifica è stata dimostrata nel 1976, quando un gruppo di ricercatori dell'Università di Birmingham, con il quale collaboro, ha lanciato mediante un razzo sonda un piccolo telescopio di questo tipo. La prima applicazione su grande scala della nuova tecnica si è avuta nel 1985, quando due telescopi affiancati costruiti dal nostro gruppo sono stati impiegati a bor-

Le maschere di codifica possono risultare estremamente diverse a seconda del particolare insieme utilizzato per ricavarle. Quelle riportate nell'illustrazione sono state ottenute annerendo le caselle che rappresentano le posizioni dei fori della maschera su griglie di 4095 (*a sinistra*), 3599 (*al centro*) e 6643 caselle (*a destra*). Considerato che i fori indicati hanno le stesse dimensioni delle caselle della griglia, nelle prime due maschere le caselle adiacenti si uniscono, formando strutture più grandi sia regolari, sia irregolari.

do della navetta statunitense nel corso della missione *Spacelab 2*.

I due telescopi utilizzavano maschere diverse, ottimizzate per compiti differenti. Il potere risolvente di un telescopio a maschera di codifica, come quello di uno stenoscopio, dipende solo dalle dimensioni dei fori e dalla loro distanza dal piano sul quale si registra l'immagine. Entrambi i telescopi erano lunghi circa tre metri (quanto consentivano le dimensioni dello spazio di carico), ma uno dei due aveva una maschera con fori di circa due millimetri di diametro, che gli consentivano di raggiungere una risoluzione di tre minuti d'arco, circa un decimo del diametro della Luna. La maschera del secondo telescopio, invece, aveva fori quattro volte più grandi e di conseguenza una risoluzione inferiore, ma risultava più sensibile a sorgenti di raggi X estese e diffuse.

Nel corso della missione, durata otto giorni, sono state riprese parecchie sorgenti di raggi X, in particolare quelle coincidenti con ammassi di galassie e con la zona centrale della nostra galassia. Periodicamente i dati ricavati da brevi osservazioni campione venivano analizzati al calcolatore presso il Lyndon B. Johnson Space Center di Houston quasi subito dopo essere stati registrati. Le rozze immagini così ottenute venivano usate per controllare il comportamento degli strumenti. A causa della natura intrinsecamente indiretta della visualizzazione mediante maschera di codifica, tuttavia, l'autentica ricchezza delle informazioni ottenute poté venire apprezzata solo dopo il termine della missione.

Il gruppo di Birmingham, collaborando con il Laboratorio di Ricerca Spaziale di Utrecht, ha costruito anche un telescopio a maschera di codifica più piccolo, che costituisce uno dei quattro strumenti per astronomia X e gamma a bordo del modulo *Kvant*, unito alla stazione spaziale sovietica *Mir*. Gli strumenti del modulo *Kvant* consentono di osservare molti oggetti e di studiarne le variazioni di luminosità. Uno degli obiettivi principali del nostro telescopio è stata la supernova nella Grande Nube di Magellano, che, per un caso fortunato, è apparsa poche settimane prima della messa in orbita del modulo. Il telescopio è stato utilizzato per rilevare la prevista emissione di raggi X da parte della supernova nel momento in cui i frammenti dell'esplosione si espandono, rivelando il materiale radioattivo del nucleo.

Le immagini riprese dai telescopi della missione *Spacelab 2* e del modulo *Kvant* hanno permesso di individuare molte sorgenti di raggi X prima sconosciute e stanno fornendo un mezzo per studiare la struttura delle emissioni X a energie alle quali ciò non era stato possibile in precedenza. Di particolare interesse è il fatto che parte dei raggi X provenga da una zona che coincide per posizione con il nucleo della nostra galassia, dove alcuni astronomi ritengono debba trovarsi un buco nero di massa estremamente elevata. Questo oggetto, con una massa un milione di volte maggiore di quella del Sole, potrebbe senz'altro dare origine ai processi di elevatissima energia che sono stati osservati in quella zona.

Malgrado i molti vantaggi dei telescopi a maschera di codifica, gli strumenti a incidenza radente sembrano preferibili per la visualizzazione di sorgenti di raggi X molli. Innanzitutto, oggi essi consentono la migliore risoluzione; inoltre sono in grado di dirigere la radiazione X proveniente da una data zona del campo di osservazione in una parte ben definita di un piccolo rivelatore. Ciò consente di ridurre a un livello molto basso il rumore di fondo dovuto alla radiazione cosmica e ai raggi X provenienti da altre parti del campo osservato.

I telescopi che utilizzano maschere di codifica dell'immagine saranno invece molto utili in tutti quei casi in cui, per un motivo o per l'altro, non si possono usare strumenti di altro tipo. La ragione più frequente è data dalla necessità di visualizzare sorgenti di raggi X duri o anche di raggi gamma molli; un'altra ragione può essere la necessità di coprire un campo visivo molto ampio, come quando si voglia rilevare una vasta zona di cielo.

Inoltre, dato che la risoluzione di un telescopio a maschera di codifica dipende dalla sua lunghezza, si può potenzialmente ottenere una risoluzione assai elevata costruendo uno strumento molto lungo. Per questo motivo si sta prendendo in considerazione la possibilità di costruire telescopi lunghi centinaia di metri, o anche con maschera e rivelatore posti su due diverse piattaforme spaziali. L'acquisizione delle tecnologie necessarie per costruire grandi strutture orbitanti potrebbe consentire in futuro la realizzazione di simili telescopi, che permetterebbero di risolvere dettagli nelle zone X e gamma dello spettro elettromagnetico con una precisione molto superiore a quella possibile oggi.

BIBLIOGRAFIA

BAUMERT LEONARD D., *Cyclic Difference Sets* in «Lecture Notes in Mathematics», 182, 1971.

FENIMORE E. E. e CANNON T. M., *Coded Aperture Imaging with Uniformly Redundant Arrays* in «Applied Optics», 17, n. 3, 1 febbraio 1978.

HARWIT MARTIN e SLOANE NEIL J. A., *Hadamard Transform Optics*, Academic Press, 1979.

X- and γ-Ray Imaging Techniques in «Nuclear Instruments & Methods in Physics Research», 221, n. 1, 15 marzo 1984.

Il calcolatore di Atanasoff

Contrariamente a quanto si pensava fino a non molti anni fa, il vero inventore del calcolatore elettronico digitale è stato John V. Atanasoff, uno scienziato le cui ricerche sono state per lungo tempo trascurate

di Allan R. Mackintosh

Finalmente la storia sta rendendo giustizia a John V. Atanasoff. Dopo decenni di oscurità, viene oggi finalmente riconosciuto a questo ottantaquattrenne professore di fisica in pensione il merito di un'impresa compiuta quasi mezzo secolo addietro: l'invenzione del primo calcolatore elettronico digitale. Fino a pochissimo tempo fa, la storia della nascita dell'informatica attribuiva generalmente questa invenzione ad altri.

Si è affermato a lungo, infatti, che i calcolatori che conosciamo oggi hanno avuto origine negli anni trenta e all'inizio degli anni quaranta, periodo in cui furono compiuti molti tentativi complementari e concorrenziali per automatizzare, accelerare o comunque eliminare l'ingrato compito dell'esecuzione di lunghi calcoli di routine. Per esempio, nel 1932 Vannevar Bush del Massachusetts Institute of Technology mise a punto un calcolatore meccanico, chiamato analizzatore differenziale, che effettuava calcoli mediante la rotazione di ruote dentate e alberi. Verso la fine degli anni trenta, Konrad Zuse in Germania, George R. Stibitz dei Bell Telephone Laboratories e Howard H. Aiken della Harvard University (in collaborazione con la International Business Machines Corporation) costruirono, indipendentemente l'uno dall'altro, calcolatori «elettromeccanici», nei quali i numeri erano rappresentati da una serie di dispositivi comandati elettricamente: i relè. Le posizioni «attivato» e «disattivato» dei relè corrispondevano alle cifre 0 e 1 nel sistema di numerazione binario, cioè in base 2. (A differenza del tradizionale sistema di numerazione decimale o in base 10, nel quale i numeri sono rappresentati con le cifre da 0 a 9, nel sistema binario i numeri sono rappresentati mediante le sole cifre 0 e 1.)

La storia dell'informatica affermava inoltre che i primi calcolatori elettronici furono inventati verso la metà degli anni quaranta. A differenza dei calcolatori meccanici o elettromeccanici, essi funzionano essenzialmente grazie a dispositivi elettronici, dalle valvole termoioniche e dai transistori fino ai microchip (micropiastrine) di oggi; a muoversi sono in questo caso gli elettroni e non le parti del calcolatore. In genere si conveniva che la prima macchina di questo tipo fosse stata il Colossus, costruito dai matematici Alan M. Turing e M. H. A. Newman e dai loro colleghi al Bletchley Research Establishment in Inghilterra, che cominciò a funzionare nel 1943. Il Colossus contribuì alla decifrazione del codice tedesco Enigma e quindi ebbe un effetto decisivo sull'esito della seconda guerra mondiale. Si riteneva che il secondo calcolatore fosse stato l'ENIAC (Electronic Numerical Integrator and Computer), costruito da John W. Mauchly e J. Presper Eckert e dai loro colleghi dell'Università della Pennsylvania e ultimato nel 1945.

In realtà fra il 1937 e il 1942, cioè ben prima che fossero concepite queste imponenti e importanti macchine, Atanasoff aveva progettato e costruito due calcolatori elettronici più piccoli. Il primo era un prototipo di una macchina su scala più grande, oggi nota con il nome di Atanasoff-Berry Computer, o ABC. Berry era il compianto Clifford E. Berry, un laureando di Atanasoff che divenne in seguito suo stretto collaboratore nel periodo dal 1939 al 1942.

Il tardivo riconoscimento dell'impresa di Atanasoff non è il risultato di ricerche specialistiche, bensì il frutto accidentale di una causa civile apertasi nel 1967 fra la Sperry Rand Corporation e la Honeywell, Inc. La Sperry aveva acquistato il brevetto dell'ENIAC e si faceva pagare i diritti dagli altri costruttori di calcolatori elettronici. La Honeywell si rifiutò di pagare tali diritti e di conseguenza la Sperry la citò in giudizio; nello stesso tempo la Honeywell citò la Sperry per violazione delle norme antitrust e per il tentativo di imporre un brevetto che non poteva avere alcun valore.

La Honeywell sosteneva che il brevetto non era valido perché, preparandosi a «dar battaglia» alla Sperry, gli avvocati dell'azienda si erano imbattuti nel nome di Atanasoff. Quando riuscirono a rintracciarlo, Atanasoff, che non era al corrente della tecnica di funzionamento dell'ENIAC, fu in grado di confrontare quella macchina con la sua. Così si rese conto che parti del brevetto ENIAC (che comprendevano in sostanza tutti gli aspetti dell'elaborazione elettronica) derivavano dall'ABC e dalle informazioni che egli aveva fornito a Mauchly all'inizio degli anni quaranta.

Molto colpito dalla testimonianza di Atanasoff, il giudice Earl R. Larson della Corte distrettuale di Minneapolis sentenziò il 19 ottobre 1973 che il brevetto dell'ENIAC era nullo. Egli ritenne che Mauchly ed Eckert «non avessero inventato per primi il calcolatore elettronico digitale automatico, ma avessero invece tratto spunto per esso da un certo dottor John Vincent Atanasoff». Sia durante il processo sia dopo, Mauchly rifiutò di ammettere di aver imparato alcunché di importante da Atanasoff. Anche la vedova di Mauchly, Eckert e altri sono del medesimo parere, ma personalmente ritengo che la deposizione in tribunale smentisca chiaramente la posizione di Mauchly.

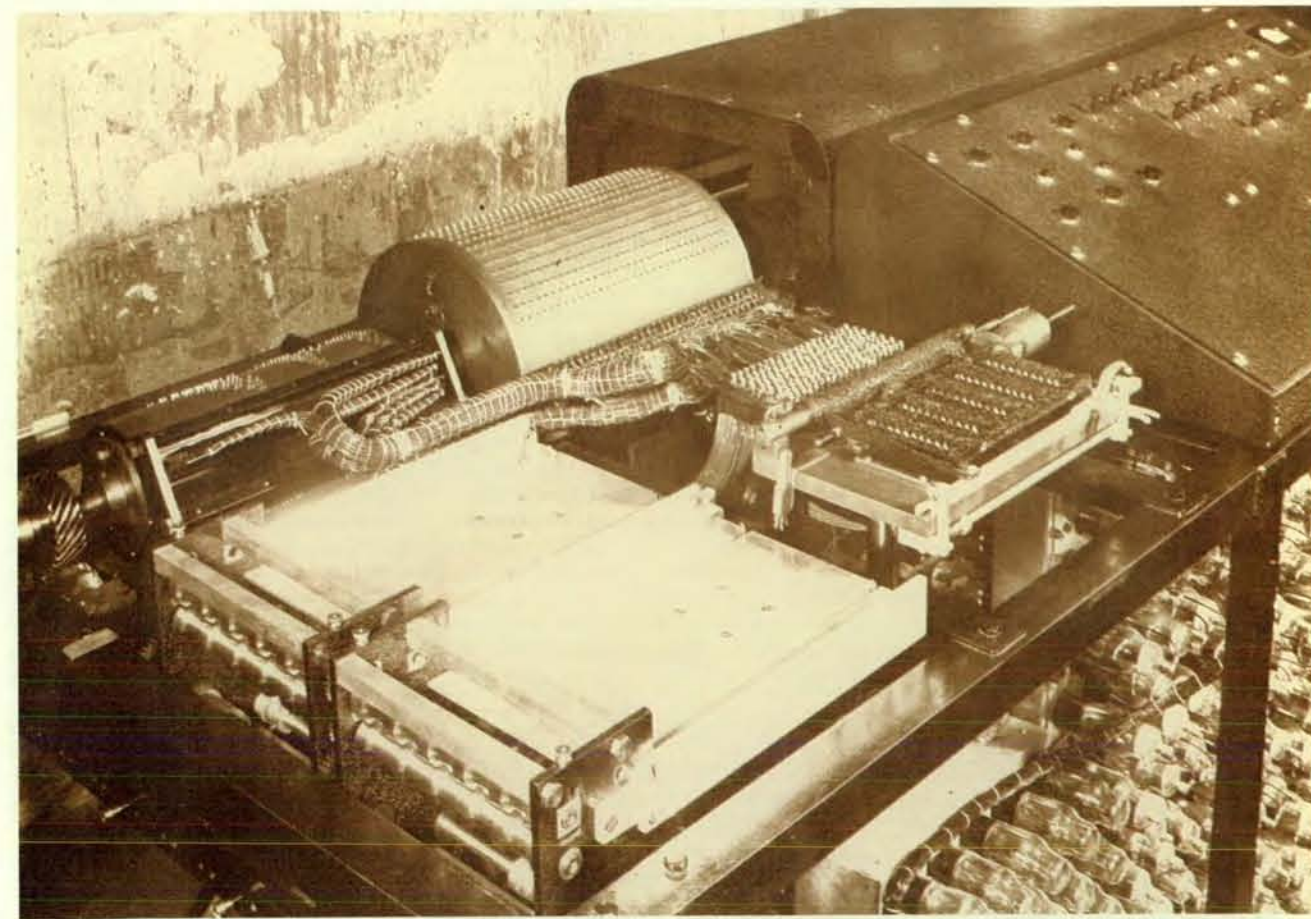
La sentenza del giudice Larson, che la Sperry accettò senza ricorrere in appello, non portò immediatamente Atanasoff alla celebrità, anche perché in quel momento, negli Stati Uniti, giornali e televisioni si stavano occupando dello scandalo Watergate che portò alle dimissioni del presidente Richard M. Nixon. Con il passare del tempo tuttavia, le notizie relative ai contributi di Atanasoff si diffusero nella comunità scientifica e il fatto che egli sia stato il primo a progettare e costruire un calcolatore elettronico digitale è oggi ampiamente riconosciuto. Gran parte del merito di questo riconoscimento va ad Arthur W. Burks, che prese parte alla costruzione dell'ENIAC, e a sua moglie Alice. I Burks, che sono rispettivamente professore e ricercatrice al Dipartimento di ingegneria elettrotecnica e informatica dell'Università del Michigan, hanno studiato minuziosamente il lavoro di Atanasoff sull'ABC e l'hanno descritto - insieme con il processo per il brevetto - in un articolo di grande risonanza e, di recente, in un libro.

La vicenda che portò alla costruzione dell'Atanasoff-Berry Computer ebbe inizio verso la fine degli anni venti, quando Atanasoff lavorava alla sua tesi di dottorato in fisica teorica all'Università del Wisconsin, a Madison. Questa tesi, sulla struttura elettronica dell'elio, lo obbligò a molte settimane di laboriosi calcoli con una calcolatrice da tavolo e gli fece rimpiangere che non esistesse un metodo per eseguirli in modo automatico. Quest'idea non lasciò Atanasoff neppure dopo che egli ebbe conseguito il dottorato, nel 1930, e fu assunto come docente allo Iowa State College (poi trasformato in Università).

Qui egli rifletté per molti anni su come realizzare l'automazione del calcolo. Nell'inverno del 1937 aveva stabilito alcuni principi generali. Per esempio, aveva deciso che la funzione di memoria, cioè l'immagazzinamento dei dati, doveva essere separata dalla funzione di elaborazione e che il metodo di calcolo doveva essere digitale anziché analogico: la macchina doveva cioè rappresentare i numeri mediante cifre e non tramite l'analogia con qualche grandezza fisica, come per esempio la distanza lungo l'asse di un regolo. Atanasoff si era anche baloccato con l'idea di usare per i calcoli una base diversa da 10. Sembrava tuttavia che queste idee non «quagliassero», come diceva lui, e ciò lo angosciava sempre più. Poi, una notte di quel cupissimo inverno, fece alcuni progressi decisivi.

All'inizio la serata non era stata molto promettente. Anzi, era stata tanto deludente che Atanasoff aveva abbandonato il laboratorio, si era messo in macchina e si era diretto ad alta velocità da Ames, dov'era la scuola, verso est, concentrandosi sulla guida per sgombrare la mente dai pensieri. Dopo alcune ore si era ritrovato a 300 chilometri di distanza, nello stato dell'Illinois, dove si fermò a bere qualcosa in un locale illuminato a giorno lungo la strada.

«Fuori era freddissimo ed entrando mi tolsi il cappotto; - ricordò durante la sua deposizione al processo - avevo un cappotto molto pesante, lo appesi e mi sedetti, poi ordinai un bicchierino e mentre aspettavo che me lo portassero, mi resi conto che non ero più tanto nervoso e che i miei pensieri tornavano a concentrarsi sulle macchine per il calcolo.

«Ebbene, non so perché la mia mente si mise a funzionare mentre prima non ci riusciva: tutto sembrava calmo e tranquillo, le cose filavano bene... Può darsi che ne abbia bevuto due, di bicchierini, perché mi accorsi che i pensieri mi flui-

Il calcolatore di Atanasoff e Berry (ABC) fu costruito fra il 1937 e il 1942 da John V. Atanasoff, allora professore di fisica allo Iowa State College (ora Iowa State University), con l'aiuto del suo laureando Clifford E. Berry. L'ABC non fu il primo calcolatore digitale progettato, poiché già in precedenza altre macchine avevano operato direttamente sui numeri anziché rappresentarli tramite una grandezza fisica, come la rotazione di un indice. Tuttavia l'ABC fu il primo calcolatore in cui dispositivi elettronici, sotto forma di valvole termoioniche, (*in basso a destra*) erano alla base delle operazioni del calcolatore e dell'aritmetica digitale. L'ABC aveva anche l'originale caratteristica di tener separate la memoria e gli elementi di elaborazione. La memoria era costituita da condensatori (dispositivi che accumulano carica elettrica) fissati a grandi tamburi (*sullo sfondo*). I «vassoi» a sinistra sono lettori di schede perforate.

Atanasoff festeggiò il suo ottantesimo compleanno alla Iowa State University nel 1983. In primo piano, un tamburo della memoria dell'ABC, l'unico componente superstite di una certa importanza. Ciascun anello di condensatori sul tamburo poteva contenere un numero binario lungo fino a 50 cifre o bit. Il tamburo conteneva 1500 bit. Oggi la memoria di un orologio calcolatore come quello che Atanasoff porta al polso può essere 10 volte più grande.

vano facilmente e che ottenevo risultati positivi.»

Positivi, eccome. Per attuare le funzioni di controllo e aritmetiche del calcolatore, Atanasoff pensò di ricorrere a commutatori elettronici (dispositivi elettronici che guidano il flusso dei segnali elettrici) anziché a quelli meccanici. In questo egli fu un antesignano: prima di allora nessuna macchina concepita per risolvere problemi matematici complessi si era basata sull'elettronica.

Atanasoff decise anche che la sua macchina digitale avrebbe operato con numeri in base 2, elaborando inoltre secondo le regole della logica anziché mediante il conteggio diretto (*si veda l'illustrazione a pagina 100*). La stessa sera risolse anche un problema particolare, che riguardava la memorizzazione dei numeri in base 2. Già in precedenza aveva considerato la possibilità di usare i condensatori, dispositivi che immagazzinano carica elettrica, per la memoria del calcolatore. Così, una carica positiva su un'armatura di un condensatore poteva rappresentare il numero 1, mentre una carica nulla poteva rappresentare lo 0. Il problema era che la carica di un condensatore tende a dissiparsi. Nella calma del locale ad Atanasoff venne l'idea di «rigenerare» la memoria con un procedimento che egli battezzò *jogging*: avrebbe ripristinato la carica del condensatore, in modo che se esso si fosse trovato, supponiamo, nello stato 1, vi sarebbe rimasto e non avrebbe cambiato stato nel tempo scendendo verso lo 0.

Presa questa decisione, ricorda Atanasoff, «a un'ora piuttosto tarda risalii in macchina e me ne tornai a casa guidando con più calma.»

Dato che i calcolatori moderni continuano a elaborare elettronicamente cifre binarie memorizzate appunto secondo le regole della logica binaria e a tenere distinta l'elaborazione dalla memoria (che si rigenera), risulta opportuno esaminare più da vicino le pionieristiche decisioni di Atanasoff. Per esempio, perché una macchina digitale è preferibile a una di tipo analogico per soddisfare le esigenze del calcolo?

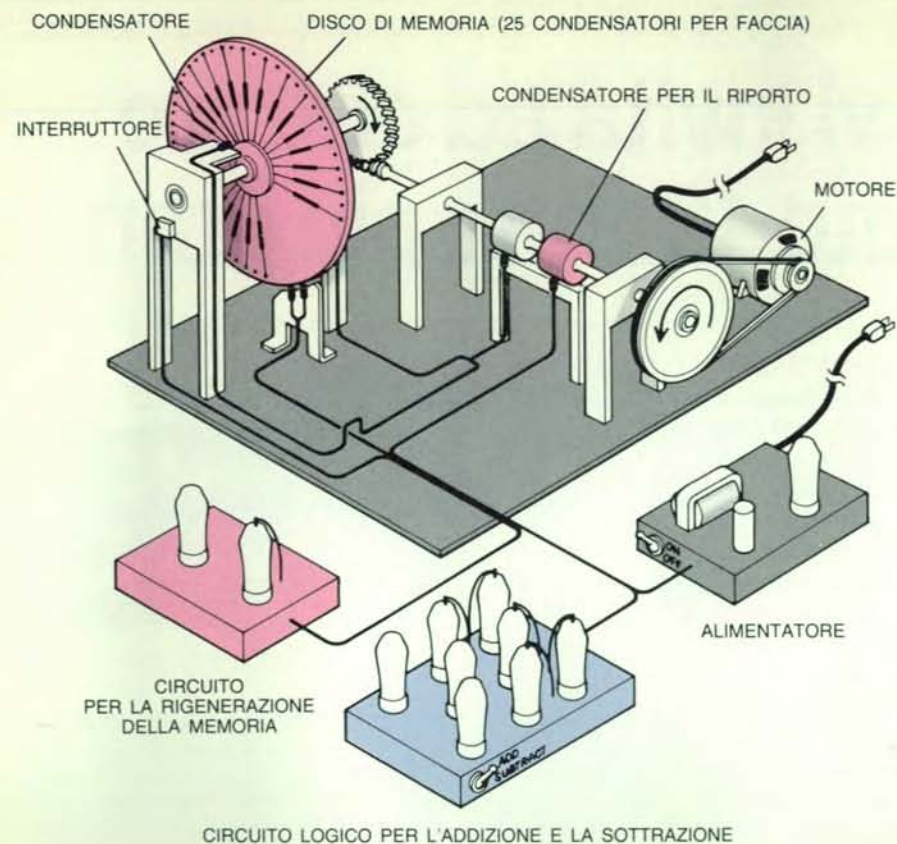
L'acutezza di Atanasoff in questo campo può essere meglio valutata se si confronta l'ABC con l'analizzatore differenziale di Bush, che era il calcolatore scientifico più progredito del tempo. Oltre a essere sostanzialmente meccanico, l'analizzatore era un calcolatore analogico: i risultati erano rappresentati mediante la rotazione di un albero.

I calcolatori analogici sono adatti a molte applicazioni, ma poiché, invece di operare su numeri, lavorano con grandezze fisiche, essi sono soggetti a una perdita inevitabile di precisione. Il calcolatore digitale di Atanasoff raggiungeva facilmente una precisione 1000 volte maggiore di quella che si poteva conseguire con l'analizzatore differenziale. Inoltre la precisione poteva essere facilmente aumentata, se necessario, aggiungendo altre cifre. Accrescere la precisione dei calcolatori analogici è difficile e costosissimo: per aumentare di un fattore 10 la precisione di un regolo calcolatore, si dovrebbe aumentare dello stesso fattore la sua lunghezza.

Oggi il calcolo digitale è basato sul sistema di numerazione binario. È evidente che Atanasoff non fu l'unico a pensare a questa soluzione (numerosi calcolatori elettromeccanici erano binari), ma egli fu il primo a escogitare un metodo elettronico per operare con cifre binarie. Che aspetto ha un numero in base 2? In base 10 ciascuna cifra di un numero rappresenta, a seconda della posizione (da destra a sinistra), un numero corrispondente di unità, decine, centinaia, migliaia e così via. Quindi il numero 237 in realtà sta per: 2 volte 10^2 più 3 volte 10^1 più 7 volte 10^0 (qualsiasi numero elevato a zero è uguale a 1). In base 2 ciascuna cifra binaria, o bit, rappresenta a seconda della posizione un numero corrispondente di 1, di 2 (2^1), di 4 (2^2), di 8 (2^3), di 16 (2^4) e così via. Quindi il numero che in base 10 si scrive 237, in base 2 si scrive 11101101; contando infatti da sinistra a destra, questo numero «contiene» un'unità di 2^7 (128 in base 10), di 2^6 (64), di 2^5 (32), di 2^3 (8), di 2^2 (4) e di 2^0 (1) e nessuna unità di 2^4 e di 2^1.

Naturalmente il sistema binario non sarebbe pratico per l'impiego quotidiano, ma dato che tutti i numeri sono rappresentati in termini di cifre 1 e 0, esso offre al programmatore il vantaggio fondamentale di poter rappresentare qualunque numero mediante un insieme di elementi che possono assumere due stati, proprio come gli stati «carico» e «scarico» dei condensatori di Atanasoff o la magnetizzazione «verso l'alto» oppure «verso il basso» delle regioni di un disco magnetico.

Atanasoff decise di immagazzinare le cifre binarie nei condensatori dopo aver esaminato diverse alternative, come le valvole termoioniche e i materiali ferromagnetici (nei quali l'orientazione di minuscoli magneti può essere modificata applicando un campo magnetico). Scelse i condensatori perché sono piuttosto economici e possono inviare i segnali all'unità di elaborazione senza bisogno che vengano amplificati. Questa scelta, come la sua soluzione di ricaricare i dispositivi di memoria, continua a condizionare l'elaborazione contemporanea. Oggi i condensatori sono parti fondamentali dei microchip che costituiscono le memorie dinamiche dei moderni calcolatori e il *jogging* di Atanasoff ha un'importanza essenziale nel funzionamento della memoria.

La scoperta di un metodo per conservare la memoria nei condensatori fu senza dubbio importante, ma probabilmente l'impresa di maggior rilievo di Atana-

Il prototipo dell'ABC fu costruito nel 1939 per collaudare due teorie fondamentali di Atanasoff. Al fine di evitare che i condensatori della memoria perdessero la carica in modo incontrollabile, Atanasoff aveva pensato di ricaricarli con continuità. Inoltre egli intendeva eseguire i calcoli mediante circuiti logici, cioè utilizzando gruppi di valvole termoioniche che sommassero o sottraessero i numeri binari in base a regole logiche anziché eseguendo conteggi. Il prototipo funzionava benissimo. A ogni giro compiuto dal disco della memoria (*ruota in rosso*), i cui condensatori contenevano su ogni faccia un numero binario di 25 cifre, il circuito logico (*al centro in basso*) sommava o sottraeva il numero registrato su una faccia del disco al numero presente sull'altra faccia. Mentre il circuito eseguiva i calcoli (andando anche a depositare e a estrarre le cifre del riporto o del prestito da un apposito condensatore), il circuito di rigenerazione (*in basso a sinistra*) ricaricava la memoria.

soff fu il progetto di un complesso commutatore elettronico detto circuito logico. Mentre si trovava in quel locale dell'Illinois, egli aveva previsto due unità di memoria, che aveva battezzato «abaci». Poi immaginò, per usare le sue parole, una «scatola nera» - il circuito logico - in cui avrebbero dovuto essere trasferiti i numeri contenuti nella memoria; sulla base delle regole logiche espresse nei collegamenti fissi, la scatola nera avrebbe poi fornito ai terminali d'uscita il risultato corretto dell'addizione o della sottrazione di questi numeri.

Egli decise di costruire la scatola nera con valvole termoioniche. Esse dovevano ricevere i segnali dai condensatori delle memorie, che egli battezzò «abaco tastiera» e «abaco contatore», per analogia rispettivamente con la tastiera e con il carrello mobile - il contatore o accumulatore - delle calcolatrici meccaniche da tavolo usate a quel tempo. Le valvole ricevevano segnali anche da altri condensatori che immagazzinavano le cifre del riporto (nel caso dell'addizione) o le cifre prese «a prestito» (nel caso della sottrazione). «Essendo stato istruito da un uomo con il saldatore», il circuito logico avrebbe allora scelto la risposta giusta sostituendo il risultato a ciò che si trovava nel contatore in quel momento. Le valvole operavano sulle informazioni con tale rapidità che potevano essere sfruttate ripetutamente per sommare o sottrarre le cifre di due numeri qualunque contenuti negli abaci. Attualmente i circuiti logici sono alloggiati in minuscoli chip, che sono molto più veloci delle valvole termoioniche, ma compiono essenzialmente le stesse funzioni previste da Atanasoff.

Che cosa è accaduto dell'altra iniziativa fondamentale di Atanasoff, cioè quella di separare la memoria dall'elaborazione? Anche qui la sua eredità è viva: nei calcolatori moderni, per esempio in un microcalcolatore da tavolo, vi sono tre elementi distinti: il sistema di ingresso-uscita, che comprende in sostanza tastiera, schermo e stampante (Atanasoff decise di far perforare l'ingresso e l'uscita su schede del tipo già in uso per i calcolatori); l'unità centrale di elaborazione, in cui vengono effettuate le operazioni di controllo e di elaborazione; e la memoria, che ha componenti interne ed esterne (dischi).

Benché Atanasoff fosse convinto di aver scoperto i principi corretti del calcolo elettronico, sapeva che per tradurli in pratica sarebbe stato necessario uno sforzo prodigioso. In questa impresa egli ebbe l'aiuto fondamentale di Berry, che era ossessionato quanto Atanasoff dal calcolo elettronico. In seguito Atanasoff raccontò che entrambi erano molto occupati, ma egli non riusciva a ricordare «neppure un caso in cui uno di noi non avesse tempo per il calcolatore; i nostri cuori erano proprio lanciati in quell'avventura».

Il loro primo passo fu quello di costruire un piccolo prototipo per collaudare gli elementi essenziali del progetto di Atanasoff: il circuito logico elettronico e la memoria binaria rigenerativa. Il compito fu portato a termine con rapidità notevole: il prototipo era in grado di funzionare già nell'ottobre del 1939.

Aveva due abaci di memoria, montati ciascuno su una faccia di un disco di plastica (bachelite). Ciascun abaco consisteva di 25 condensatori ed era quindi in grado di contenere un numero binario di 25 cifre, che equivale a un numero decimale di otto cifre. Atanasoff e Berry introducevano negli abaci i numeri binari caricando a mano i condensatori che rappresentavano le cifre 1 e lasciando scarichi quelli che corrispondevano agli zeri. La pressione di un interruttore faceva compiere al disco una rotazione, durante la quale il circuito logico elettronico, formato da otto valvole termoioniche, leggeva i numeri dagli abaci. Con l'ausilio di un condensatore che conteneva le cifre del riporto, il circuito sommava allora i numeri e inseriva la risposta nell'abaco contatore, che poteva essere letto manualmente. Nello stesso tempo il numero contenuto nell'abaco tastiera veniva rigenerato da un apposito circuito.

Questo prototipo non era certo un calcolatore efficiente: con carta e matita i calcoli si eseguivano più in fretta. Eppure esso sta al calcolo elettronico come l'aeroplano dei fratelli Wright sta all'aeronautica. Dimostrando la realizzabilità dei principi di Atanasoff, il prototipo aprì il cammino che condusse al calcolatore moderno.

A questo punto Atanasoff era pronto per costruire il vero e proprio ABC, cosa che fece tra il 1939 e il 1942. L'ABC fu progettato per eseguire un compito particolare ma impegnativo, di frequente evenienza in ingegneria e in fisica, cioè la risoluzione di grandi sistemi di equa-

zioni lineari. Un sistema di due equazioni è dato per esempio dalla coppia $2x + 5y = 9$ e $x + 2y = 4$, dove x e y sono le variabili o incognite. Chiamiamo a la prima equazione e b la seconda.

Come sa chiunque abbia studiato algebra, i sistemi di equazioni lineari nelle stesse variabili possono essere risolti facilmente con il metodo di eliminazione di Gauss: una delle due equazioni viene sommata o sottratta all'altra fino a quando il coefficiente di una delle variabili si annulla ed essa scompare. Nell'esempio considerato, se si sottrae due volte b da a il coefficiente 2 di $2x$ si riduce a 0 e quindi si ottiene l'equazione $y = 1$. Sostituendo 1 a y nell'equazione a di partenza si ottiene $x = 2$. Si osservi che sottrarre b da a due volte equivale a moltiplicare b per 2 e poi sottrarla da a; la moltiplicazione, in fin dei conti, è solo un'addizione ripetuta.

Atanasoff naturalmente aveva preso in considerazione un problema più complesso: voleva risolvere n equazioni in n incognite e precisamente un sistema di 29 equazioni nelle 29 incognite $x_1, \ldots, x_{29}$. La risoluzione di questi sistemi di equazioni ricalca lo schema precedente. Come nel caso descritto, si considerano due equazioni, per fare un esempio $2x_1 + 5x_2 - 3x_3 + 7x_4 + \ldots + 6x_{29} = 9$ e $x_1 + 2x_2 + 4x_3 - 2x_4 + \ldots + 8x_{29} = 4$, e si sottrae un multiplo dell'una dall'altra in modo da eliminare una delle incognite. Per esempio, per eliminare x_1 risulta necessario moltiplicare la seconda equazione per 2 e poi sottrarla alla prima, il che fornisce la seguente equazione: $x_2 - 11x_3 + 11x_4 + \ldots - 10x_{29} = 1$.

Applicando questo procedimento a diverse coppie di equazioni, si possono ottenere 28 equazioni dalle quali è stata eliminata la variabile x_1. Ripetendo il procedimento su queste 28 equazioni si ottengono 27 equazioni nelle quali non compaiono né x_1 né x_2, e così si prosegue fino a quando resta una sola equazione in una sola incognita. A questo punto è piuttosto facile ripercorrere il cammino a ritroso, determinando il valore di tutte le incognite.

Il metodo è elementare, ma è evidente che comporta una quantità enorme di calcoli. Atanasoff stimò realisticamente che per risolvere un sistema di 29 equazioni con una calcolatrice da tavolo di vecchio tipo sarebbero occorse circa 10 settimane di sforzi tediosissimi, mentre il suo calcolatore avrebbe svolto il compito in una o due settimane.

Per riuscire a risolvere sistemi di molte equazioni, Atanasoff collocò l'abaco tastiera e l'abaco contatore dell'ABC su grandi tamburi anziché su un disco. Ciascun tamburo poteva contenere 30 numeri binari, fino a un massimo di 50 cifre ciascuno.

L'ABC eseguiva il procedimento di eliminazione di Gauss su due equazioni alla volta. I coefficienti delle equazioni, che in precedenza erano stati perforati in forma decimale su schede, venivano trasformati in base 2 da un apposito tamburo di conversione e immagazzinati in memoria. I coefficienti di una delle equazioni venivano caricati nel tamburo con gli abaci contatori e quelli dell'altra andavano al tamburo con gli abaci tastiera. A ogni rotazione dei tamburi, che richiedeva un secondo, i circuiti logici eseguivano un'addizione o una sottrazione sui due gruppi di coefficienti. Più precisamente, un circuito logico, che consisteva in sette valvole, sommava o sottraeva il coefficiente di x_1 che si trovava nell'abaco tastiera al coefficiente di x_1 che si trovava nell'abaco contatore, lasciando poi la somma o la differenza in quest'ultimo. Al contempo gli altri circuiti operavano in modo analogo sulle altre coppie di coefficienti. (Questo procedimento, nel quale parecchie operazioni identiche vengono eseguite in parallelo, configura un'operazione vettoriale, sicché una macchina che effettui operazioni del genere è detta calcolatore vettoriale.) Nel frattempo altri circuiti ricaricavano l'abaco tastiera, rigenerando la memoria.

Dopo aver eseguito parecchie di queste sottrazioni e addizioni e dopo aver eliminato un determinato coefficiente, l'ABC perforava su schede, in forma binaria, l'insieme dei coefficienti restanti. Le schede venivano quindi messe da parte, pronte per essere utilizzate in un passo successivo, quando un lettore di schede binarie trasferiva le loro informazioni nella memoria. Una volta che i valori di tutte le variabili erano stati ottenuti in forma binaria, il lettore di schede decimali effettuava l'operazione inversa, traducendo nuovamente i dati binari in numeri decimali.

Il sistema di ingresso e uscita a schede perforate aveva funzionato bene nei collaudi preliminari, ma quando fu inserito nell'ABC esso commetteva un errore ogni 10 000 operazioni circa di perforazione e di lettura. Di conseguenza, un sistema di molte equazioni non poteva essere trattato in modo soddisfacente, cioè senza numerosi controlli e ripetizioni dei calcoli, anche se i sistemi piccoli potevano essere risolti facilmente. Atanasoff e Berry stavano ancora cercando di risolvere questo problema piuttosto banale quando la seconda guerra mondiale li costrinse a interrompere bruscamente il loro lavoro. Berry chiese il rinvio del servizio militare e Atanasoff entrò a far parte del Naval Ordnance Laboratory degli Stati Uniti.

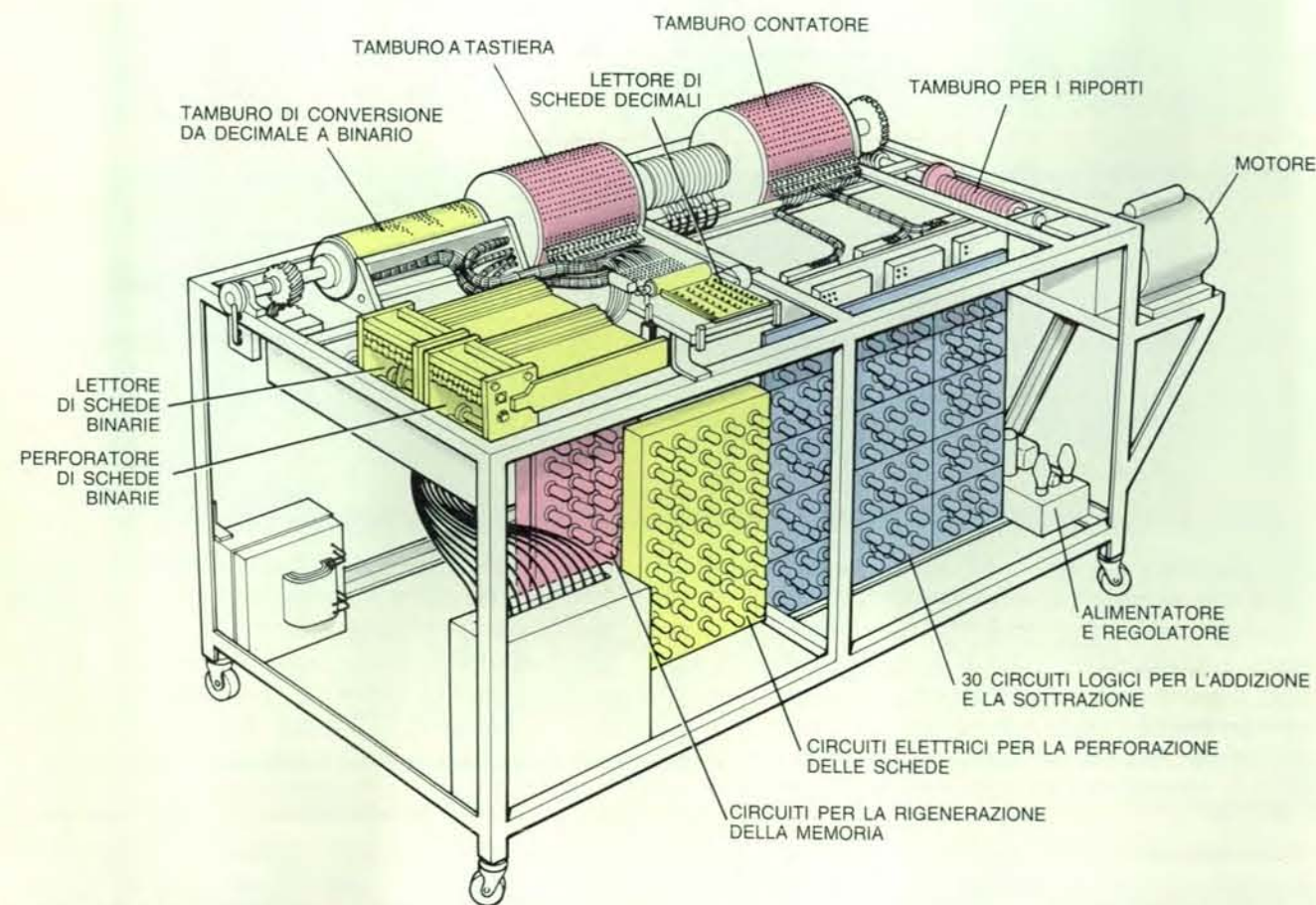
Oggi il calcolatore che essi dovettero abbandonare è spesso descritto come una macchina incompiuta. Sarebbe più esatto dire che era un calcolatore funzionante ma fallibile, in cui la parte elettronica adibita al calcolo, cioè il complesso dei circuiti logici, costituiva una brillante realizzazione. Considerando la grande rapidità con cui l'ABC fu progettato e costruito, si può quasi certamente presumere che il problema posto dal sistema di conversione delle schede binarie sarebbe stato risolto piuttosto rapidamente. In effetti sarebbe stato adatto allo scopo un sistema di ingresso e uscita costruito decenni prima dalla IBM (e in seguito incorporato nell'ENIAC). Inoltre, con la dimostrazione delle capacità del suo calcolatore, Atanasoff avrebbe certo potuto ottenere una sovvenzione per portare a termine un progetto così importante.

Se Atanasoff e Berry avessero potuto continuare, non c'è dubbio che l'ABC sarebbe stato pienamente funzionante nel 1943. Invece subì la sorte di gran parte dei dispositivi che vengono abbandonati a invecchiare: dapprima gli furono sottratte varie parti e alla fine fu smantellato all'insaputa di Atanasoff.

Se l'ABC fu dimenticato per tanto tempo, com'è possibile che le idee di Atanasoff abbiano influito sull'informatica moderna? La risposta ovviamente chiama in causa Mauchly, che introdusse nell'ENIAC le innovazioni di John V. Atanasoff.

L'ENIAC era notevolmente diverso dall'ABC: era il primo calcolatore elettronico di impiego generale, mentre l'ABC era stato progettato come macchina specializzata. (L'ENIAC poteva essere programmato per la risoluzione di problemi diversi modificando la configurazione dei conduttori innestati in un pannello di controllo.) La macchina di Mauchly ed Eckert era molto più grande di quella di Atanasoff, aveva migliaia e non centinaia di valvole termoioniche ed era molto più veloce perché la sua memoria era elettronica e non ricorreva a tamburi rotanti. Inoltre l'ENIAC calcolava contando direttamente e non tramite la logica, e calcolava in base 10.

Nondimeno è evidente che Mauchly ed Eckert introdussero nell'ENIAC e in un calcolatore successivo, l'EDVAC, gli elementi fondamentali dell'elaborazione elettronica digitale di Atanasoff. Ovviamente sia l'ENIAC sia l'EDVAC ricorrevano alla commutazione elettronica per regolare il funzionamento del calcolatore; l'EDVAC inoltre impiegava circuiti logici per le operazioni aritmetiche, che venivano eseguite in base 2, e usava una memoria che si rigenerava. Mauchly mutuò da Atanasoff anche l'idea che l'elettronica digitale avrebbe consentito di costruire una macchina capace di eseguire i calcoli più rapidamente e con maggior precisione rispetto all'analizzatore differenziale di Bush.

Atanasoff, che già nel maggio 1941 sapeva «di poter costruire una macchina in grado di fare quasi qualunque cosa nell'ambito dell'elaborazione», comprese che l'ABC poteva essere trasformato in un analizzatore differenziale elettronico digitale quando venne a sapere da un collega che alcuni ricercatori del Massachusetts Institute of Technology stavano meditando di introdurre l'elettronica in una nuova versione analogica dell'analizzatore. Atanasoff scrisse a Mauchly di questa possibilità e i due ne discussero a lungo durante una visita di quasi una set-

I componenti dell'ABC di Atanasoff erano progettati per consentire alla macchina di risolvere un sistema di 29 equazioni nelle 29 variabili $x_1, \ldots, x_{29}$. Un sistema del genere si può risolvere sommando oppure sottraendo ripetutamente un'equazione a un'altra fino a quando dalla seconda equazione risulti eliminata una variabile. Per ottenere la soluzione, cioè i valori delle variabili, questo procedimento viene ripetuto numerose volte. Per eseguire questi calcoli, l'ABC leggeva i coefficienti delle variabili (per esempio il 2 nel termine $2x_1$) da schede perforate, li trasformava in base 2 e caricava i coefficienti di un'equazione nel tamburo di memoria «a tastiera» e i coefficienti dell'altra equazione nel tamburo «contatore». A ogni rotazione dei tamburi, ciascun circuito logico (formato da sette valvole termoioniche) sommava o sottraeva una coppia di coefficienti, depositando il risultato nel contatore. Nello stesso tempo i circuiti di rigenerazione della memoria ricaricavano i condensatori della tastiera. Quando l'ABC eliminava una determinata variabile, la macchina registrava i restanti coefficienti dell'equazione su schede perforate in attesa di impiegarli successivamente.

timana che Mauchly fece ad Atanasoff nel giugno 1941. Nel corso di questa visita Atanasoff gli mostrò anche il funzionamento dell'ABC, che era ormai quasi ultimato. Quattro anni dopo, le idee di Atanasoff presero corpo nell'ENIAC.

L'ENIAC e il Colossus, anch'esso programmabile, spianarono la strada per il passo successivo nell'evoluzione del calcolatore elettronico: l'inserimento di un programma nella memoria. Nelle macchine di impiego generale questo progresso non solo rese più facile e flessibile la programmazione, ma consentì anche al programma di operare in modo diverso a seconda dei risultati dei vari passi intermedi.

Da quando, verso la fine degli anni quaranta, furono introdotti i primi calcolatori con programma incorporato, queste macchine sono diventate più veloci e potenti, ma la loro architettura non si è modificata in modo decisivo. Inoltre gli usi cui vengono adibiti oggi alcuni calcolatori riecheggiano il passato. Per esempio c'è stato di recente un ritorno d'interesse per i calcolatori dedicati simili all'ABC, in particolare fra gli scienziati che devono risolvere problemi particolari. Anzi, l'ABC e un moderno elaboratore vettoriale per la risoluzione di equazioni lineari presentano somiglianze sorprendenti (sebbene le macchine odierne offrano prestazioni immensamente più veloci).

I meriti di Atanasoff sarebbero stati senza dubbio riconosciuti prima se egli avesse ottenuto un brevetto per la sua invenzione. Come sottolineano i coniugi Burks, avrebbe potuto vantare un diritto di priorità per il concetto di calcolo elettronico digitale, oltre che per l'uso della commutazione elettronica nei calcolatori, per i circuiti di addizione e sottrazione logica, per la separazione dell'elaborazione dalla memoria, per le memorie a tamburi di condensatori, per la rigenerazione della memoria, per l'impiego del sistema binario di numerazione nel calcolo elettronico, per le unità modulari, per l'elaborazione vettoriale e per il controllo a marcatempo delle operazioni elettroniche e per un gran numero di altre innovazioni.

Affermare che si sarebbe trattato di uno dei brevetti più importanti mai concessi è ancora poco. Purtroppo, a causa della confusione creata dalla guerra e dell'inettitudine delle persone incaricate di ottenere il brevetto, per nessuna delle innovazioni di Atanasoff fu mai fatta richiesta in questo senso. Da parte sua, Atanasoff non cercò di ottenere il brevetto neppure a guerra finita perché fu portato a credere che i principi di funzionamento dell'ENIAC fossero molto diversi da quelli dell'ABC e che l'ENIAC sarebbe stato il modello dei calcolatori del futuro, il che avrebbe privato di ogni valore un brevetto che proteggesse i concetti e i dispositivi dell'ABC. Inoltre egli si trovò dapprima molto impegnato in altri progetti e in seguito si dedicò alla fondazione di una società di ricerche in ingegneria.

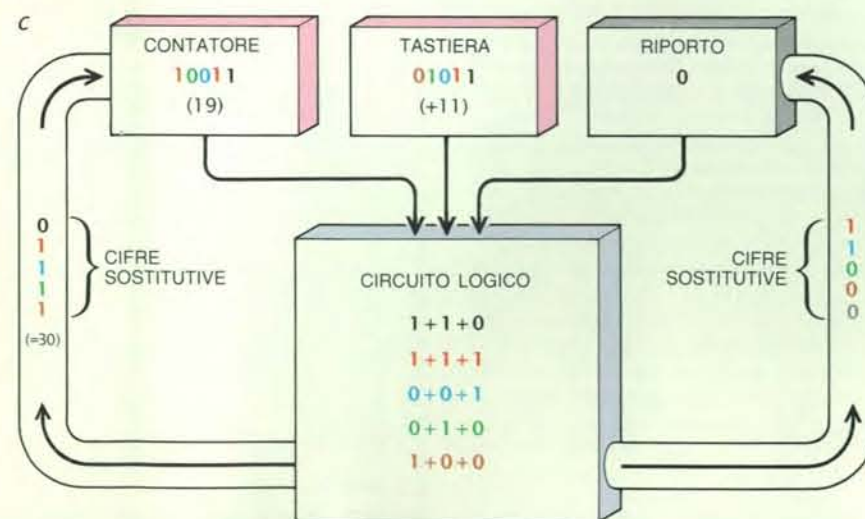
Oltre a far luce su un'innovazione tecnologica fondamentale, la storia di John V. Atanasoff offre lo spunto per compiere alcune riflessioni sull'impresa scientifica. Per prima cosa, la vita dell'inventore non è sempre facile. Nonostante i suoi grandi sforzi, Atanasoff riuscì a reperire per l'ABC soltanto 6000 dollari di sovvenzione, mentre l'ENIAC ottenne uno stanziamento di mezzo milione di dollari a causa delle sue possibili applicazioni in campo militare (tra le altre cose, esso produsse tavole di puntamento da utilizzare in artiglieria).

a

	2^4 (16)	2^3 (8)	2^2 (4)	2^1 (2)	2^0 (1)	
CONTATORE (I)	1	0	0	1	1	(19)
TASTIERA	0	1	0	1	1	(+11)
CONTATORE (U)	1	1	1	1	0	(=30)
RIPORTO	0	0	0	1	1	0

b

INGRESSO			USCITA	
CONTATORE	TASTIERA	RIPORTO	CONTATORE	RIPORTO
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	0	1
1	0	0	1	0
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1

c

Ciascun circuito logico dell'ABC sommava due numeri alla volta (*a*), in base alle regole espresse in una tabella (*b*). Si supponga di dover sommare l'equivalente binario del numero 19 contenuto nel contatore e l'equivalente binario del numero 11 contenuto nella tastiera. In base 2, ossia espressi mediante potenze di 2 ($2^0 = 1, 2^1 = 2, 2^2 = 4$...), i numeri 19 e 11 assumono rispettivamente la forma 10011 (16 + 0 + 0 + 2 + 1) e 01011 (0 + 8 + 0 + 2 + 1). Il circuito logico agiva su questi numeri sommando dapprima le cifre della colonna di destra (2^0). A questo scopo esso verificava che le cifre contenute rispettivamente nel contatore, nella tastiera e nella memoria di riporto, cioè 1,1,0, coincidessero con quelle di una riga della parte «ingresso» della tabella (in questo caso la penultima). In base alla corrispondente parte «uscita», il circuito mandava al contatore (*c*), dove venivano registrati i risultati, uno 0 (*in nero*) che sostituiva l'1 iniziale. Inoltre veniva inviato alla memoria di riporto un 1 (*in rosso*). Questo procedimento equivale a stabilire contando che 1 più 1 fa 2, numero che in notazione binaria è espresso come 10. Il circuito sommava allo stesso modo le cifre delle colonne successive fino a ottenere il risultato finale.

Un'ulteriore considerazione riguarda la creatività degli scienziati. I progressi compiuti da Atanasoff quella notte d'inverno del 1937 gettano molta luce sul processo creativo. Egli si dedicò al suo progetto immergendosi in tutti gli aspetti del calcolo automatico. Si cimentò a lungo con il problema, ricavandone molte delusioni e facendo pochi progressi evidenti, ma la sua mente continuò ad assorbire informazioni e a lavorare su di esse, in larga misura anche inconsciamente. Poi, mentre era impegnato in un'attività completamente diversa, gli arrivò la soluzione.

Al profano la corsa di 300 chilometri che portò Atanasoff fino a quel locale potrebbe apparire un modo assai poco economico per farsi un bicchiere, ma egli sapeva benissimo ciò che stava facendo. Capiva che la mente ha bisogno di svago e di riposo per essere creativa. Dopo aver concepito alcuni principi fondamentali, lasciò entrare in scena una «sorta di cognizione». Questa fiducia nell'intuizione non è forse in armonia con l'idea che si ha comunemente della ricerca scientifica, che è considerata un'attività strettamente razionale; nondimeno si tratta di un'impostazione seguita da parecchi ricercatori.

Da ultimo, non è una coincidenza che molti progressi fondamentali della tecnologia siano dovuti a scienziati: l'impegno della ricerca spesso richiede l'invenzione di nuovi strumenti e i ricercatori, che sono profondamente immersi nella risoluzione di problemi scientifici, hanno le ragioni più valide per accettare la sfida. Questo legame fra scienza e tecnologia dovrebbe essere compreso da quanti ritengono che sia possibile tagliare i fondi alla ricerca teorica senza che il progresso della tecnica ne venga in qualche modo rallentato.

BIBLIOGRAFIA

BURKS ARTHUR W. e BURKS ALICE R., *The ENIAC: First General-Purpose Electronic Computer* in «Annals of the History of Computing», 3, n. 4, ottobre 1981.

ATANASOFF JOHN VINCENT, *Advent of Electronic Digital Computing* in «Annals of the History of Computing», 6, n. 3, luglio 1984.

BURKS ALICE R. e BURKS ARTHUR W., *The First Electronic Computer: The Atanasoff Story*, University of Michigan Press, 1988.

MOLLENHOFF CLARK R., *Atanasoff: Forgotten Father of the Computer*, Iowa State University Press, 1988.

MACKINTOSH ALLAN R., *The First Electronic Computer* in «Physics Today», 40, n. 3, marzo 1988.

(RI)CREAZIONI AL CALCOLATORE

di A. K. Dewdney

La macchina mescolatrice produce onde

Gli automi cellulari, modelli al calcolatore basati su matrici di cellule che possono assumere diversi valori, si sono diffusi come un'onda nella fisica, nella matematica e in altre scienze. Ora c'è un nuovo automa cellulare che, letteralmente, crea esso stesso onde. Chiamato «macchina mescolatrice» dai suoi progettisti, imita reazioni chimiche con una precisione che raramente si è vista in altri modelli.

Le reazioni simulate dalla macchina mescolatrice hanno luogo in mezzi chimici eccitabili: due o più composti che si possono dissociare e ricombinare in presenza di un catalizzatore. Se gli stati chimici dei reagenti hanno colori differenti, si possono osservare strutture a forma di onde che si propagano lungo frontiere semplici o intricate alla perenne ricerca di un equilibrio sfuggente.

L'automa può darci una spiegazione fisica adeguata delle onde che si osservano nelle vere reazioni? Su questo problema sono attualmente impegnati i creatori della macchina mescolatrice, Martin Gerhardt e Heike Schuster dell'Università di Bielefeld, in Germania Occidentale, insieme a un numero sempre maggiore di colleghi appartenenti ad altre università.

Un automa cellulare può essere visto come una griglia infinita di cellule quadrate che avanzano passo passo a «tic» successivi, istanti discreti scanditi da un orologio immaginario. A ogni dato istante ciascuna cellula si trova in uno stato appartenente a un insieme finito di possibili stati. Lo stato di una cellula all'istante $t + 1$ dipende in modo molto semplice dallo stato in cui si trovano, all'istante t, le cellule nelle sue immediate adiacenze. Questa dipendenza è espressa da un insieme di regole che si applicano allo stesso modo a tutte le cellule della griglia. Applicando le regole a ogni istante dell'orologio, si può farcambiare ed evolvere nel tempo una qualsivoglia configurazione iniziale di stati. In alcuni casi si sviluppano forme straordinarie che spingono gli osservatori a ritenere che, data la giusta configurazione iniziale, un automa cellulare possa produrre qualcosa in grado di darsi un'organizzazione, di crescere e di riprodursi - in breve, qualcosa di «vivo».

L'automa cellulare più noto ai lettori è probabilmente il famoso gioco Vita, inventato negli anni sessanta dal matematico John Horton Conway dell'Università di Cambridge. In Vita ciascuna cellula ha solo due stati possibili: viva e morta. Le regole del gioco sono molto semplici. Se una cellula è morta all'istante t, diventerà viva all'istante $t + 1$ se esattamente tre dei suoi adiacenti sono vivi all'istante t. Se una cellula è viva all'istante t, morirà all'istante $t + 1$ se meno di due o più di tre dei suoi adiacenti sono vivi all'istante t. Bastano queste due regole perché l'automa cellulare di Vita esibisca una sorprendente varietà di comportamenti, interamente dipendenti dalla configurazione di cellule vive e morte da cui si è partiti (si vedano le *(Ri)creazioni al calcolatore* in «Le Scienze» n. 203, luglio 1985 e n. 224, aprile 1987).

La macchina mescolatrice non è uno, ma molti automi cellulari. Si sceglie una particolare versione specificando diversi parametri, per esempio il numero degli stati. Se ci sono $n + 1$ stati, ciascuno dei possibili stati di una cellula può essere rappresentato da un numero compreso tra 0 e n. Gerhardt e Schuster, per descrivere gli stati delle cellule nella loro macchina, adottano un'estensione della metafora di Conway. Una cellula nello stato 0 è detta «sana» e una nello stato n è detta «malata». Tutti gli stati intermedi hanno un grado di «contagio» che corrisponde al loro numero di stato: più questo si avvicina a n, più la cellula risulta contagiata. La macchina mescolatrice applica selettivamente a ciascuna cellula una fra tre possibili regole, in funzione dello stato in cui si trova, cioè a seconda che la cellula in questione sia sana, contagiata o malata.

Se in un dato istante la cellula è sana (cioè nello stato 0), all'istante successivo avrà un nuovo stato che dipende dal numero di cellule contagiate, A, e dal numero di cellule malate, B, che si trovano in quel momento nelle sue adiacenze, e da due parametri indicati con $k1$ e $k2$. Per essere più precisi, lo stato della cellula nell'istante $t + 1$ è dato dalla formula seguente:

$$[A/k1] + [B/k2].$$

Le parentesi quadre indicano un arrotondamento per difetto delle frazioni che racchiudono. Se, per esempio, $A/k1$ risulta uguale a 2,725, le parentesi quadre riducono quel numero a 2. Se per caso la formula ha come risultato 0, la cellula ovviamente rimarrà sana, almeno per il momento.

Se la cellula è contagiata, le sue condizioni in genere vanno via via peggiorando. Il suo stato nell'istante $t + 1$ è la somma di due numeri: il grado di contagio nelle adiacenze della cellula nell'istante t e una costante, g, che stabilizza la velocità con cui l'infezione tende a diffondersi tra le cellule. Il grado di contagio è dato dalla divisione del numero S, la somma dei numeri di stato della cellula e dei suoi adiacenti, per A, il numero di adiacenti contagiati. Una cellula che all'istante t si trova in uno stato di contagio assume quindi, all'istante $t + 1$, uno stato dato dalla formula

$$[S/A] + g.$$

La cellula contagiata, però, non può diventare «più malata» di n. Nel caso in cui il numero dato dalla formula superi n, si prende n come nuovo stato della cellula.

Infine, se la cellula è malata (nello stato n) all'istante t, diventa miracolosamente sana (assume lo stato 0) in $t + 1$.

Oltre a queste tre regole, è necessaria una definizione di che cosa si intende per «adiacenti». Storicamente si sono usati due concetti di adiacenza negli automi cellulari: l'intorno di von Neumann e l'intorno di Moore. Per von Neumann, l'intorno di una certa cellula (cioè l'insieme delle sue adiacenti) è costituito dalle quattro cellule che con essa hanno lati in comune. L'intorno di Moore comprende invece, oltre alle cellule dell'intorno di von Neumann, anche le quattro che toccano solo i vertici della cellula data - per un totale di otto cellule. Date le tre regole e la definizione di adiacenti di una cellula, per definire completamente l'automa cellulare di Gerhardt--Schuster bisogna specificare il valore di quattro parametri: n, il numero degli stati meno 1; $k1$ e $k2$, i parametri di «peso» per le cellule sane, e g, la velocità del contagio.

Un esperimento campione effettuato da Gerhardt e Schuster su una griglia 20 per 20, usando il concetto di intorno di von Neumann, rivela il comportamento tipico delle macchine mescolatrici. (Le cellule ai margini della griglia seguono le stesse regole che valgono negli altri punti dell'automa cellulare; hanno semplicemente meno cellule nel loro intorno.) Per i parametri n, $k1$ e $k2$ erano stati fissati, rispettivamente, i valori 100, 2 e 3. Quattro tipi di comportamento emersero in corrispondenza di diversi valori del parametro g. In una prova tipica, Gerhardt e Schuster davano alle 400 cellule della griglia 20 per 20 una configurazione iniziale di stati casuale, specificavano un valore di g e lasciavano effettuare alla macchina mescolatrice 10 000 cicli di calcolo. Poiché è più facile analizzare dati a una dimensione che immagini bidimensionali, Gerhardt e Schuster registravano solo il numero di cellule contagiate in ciascun ciclo, per presentare i loro risultati sotto forma di grafici come quelli che si vedono nella pagina successiva.

A valori bassi di g, non accadeva un granché. A parte qualche fluttuazione iniziale, l'attività tra le cellule tendeva a esaurirsi: le cellule finivano col godere di una monotona e immutabile salute. Con il crescere di g, invece, cominciavano ad accadere strane cose. Tanto per iniziare, la maggior parte delle cellule risultava contagiata e lo rimaneva, anche se a tratti apparivano, in modo irregolare e casuale, delle cellule sane. Gerhardt e Schuster chiamavano Tipo 1 questo particolare comportamento.

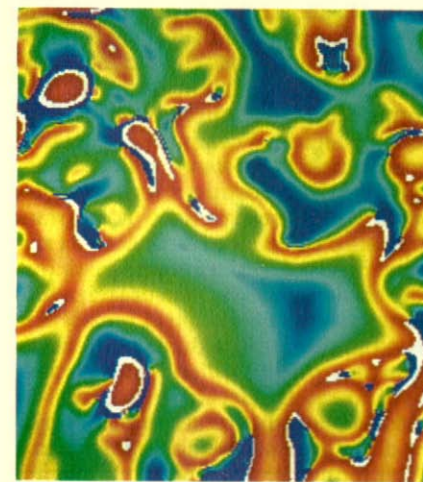
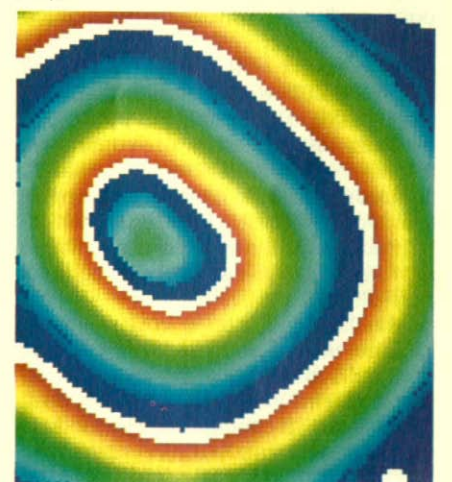
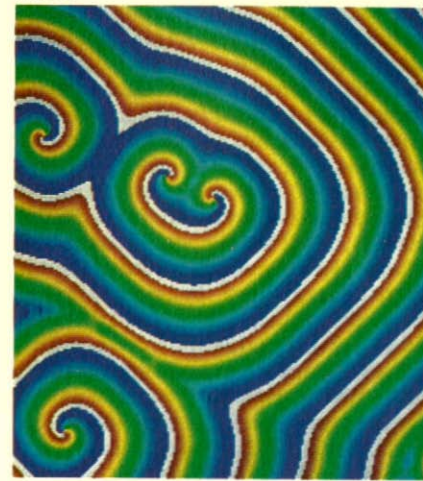
Veniva poi il Tipo 2 di comportamento, caratterizzato da una serie generalmente regolare di «epidemie» lunghe circa 30 cicli, inframmezzate dalla comparsa di un gran numero di cellule sane. (A volte quasi tutte le 400 cellule diventavano sane per poi cadere vittime di una nuova ondata di contagio.) Con l'ulteriore aumento di g, appariva il comportamento di Tipo 3, annunciato dall'instaurarsi di un'alternanza molto regolare tra saturazione e quasi totale scomparsa di cellule sane pressappoco ogni 20 cicli. Infine emergeva il comportamento di Tipo 4: dopo una brusca crescita in pochi cicli, il numero di cellule contagiate fluttuava con una certa regolarità intorno a un valore di saturazione di circa il 75 per cento.

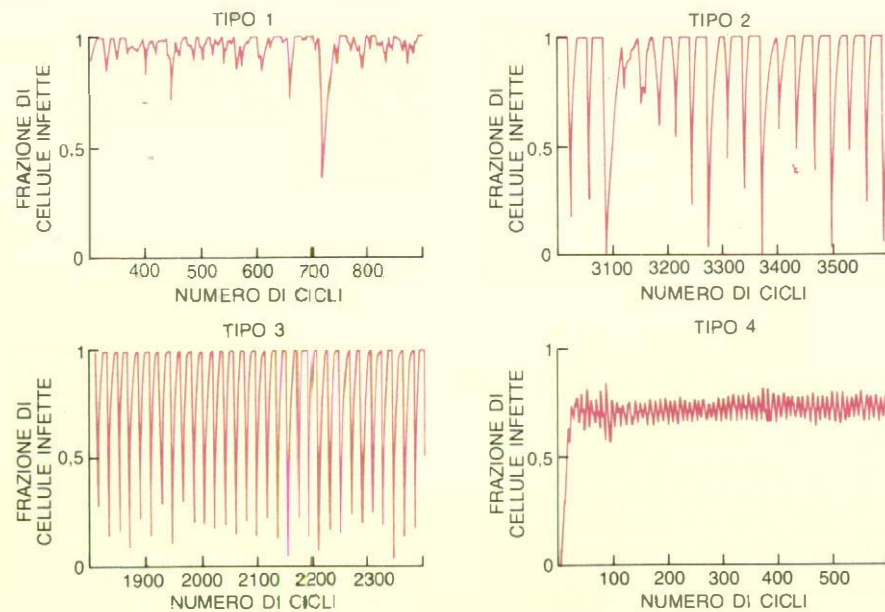
Con il progressivo crescere di g, i quattro tipi di comportamento si presentavano uno di seguito all'altro, ma con qualche sovrapposizione: per valori di g di transizione, a volte si aveva un tipo di comportamento a volte un tipo diverso. In alcuni casi, Gerhardt e Schuster furono anche testimoni di transizioni tra diversi tipi di comportamento in un'unica prova.

I quattro tipi di comportamento rappresentano la comparsa di particolari tipi di configurazioni a onda che si vedono illustrate nella figura di questa pagina. In queste immagini, le dimensioni della griglia variano da 100 per 100 cellule a 500 per 500 cellule. Le onde associate al comportamento di Tipo 1 coprivano solo brevi distanze prima di morire. Quelle di Tipo 2 si muovevano verso l'esterno in strisce circolari di ampiezza molto variabile. Quelle di Tipo 3 mostravano la stessa forma circolare ma con una maggiore regolarità, in accordo con l'andamento regolare che le cellule contagiate mostravano nel grafico. Infine, le onde di Tipo 4 seguivano una configurazione a spirale che partiva dal centro della griglia. Come sempre, i lettori che dispongono di un calcolatore sono invitati a ripetere in qualche forma l'esperimento. Le onde sullo schermo saranno sicuramente accompagnate da onde di pensiero.

Alcune delle forme ondulatorie generate dalla macchina mescolatrice sono simili a quelle mostrate da una grande varietà di sistemi chimici; alcune, in particolare, sono il ritratto perfetto delle onde chimiche scoperte nella famosa reazione di Belousov-Žabotinsky. Si confronti, per esempio, la complessa configurazione a riccioli nell'immagine generata dal calcolatore con la fotografia della reazione di Belousov-Žabotinsky nell'illustrazione di pagina 109.

A che cosa è dovuta questa somiglianza? Gerhardt e Schuster non ne rimasero del tutto sorpresi, visto che avevano deliberatamente progettato la macchina mescolatrice per imitare le caratteristiche di un particolare tipo di «reazione catalitica eterogenea» in cui monossido di carbonio e ossigeno si combinano per dare anidride carbonica, mentre sono adsorbiti sulla superficie di migliaia di piccoli cristalliti di palladio dispersi in un mezzo poroso. Al procedere della reazione di ossidazione, il calore liberato modifica lo stato del catalizzatore. Una repentina transizione di fase nel cristallite libera il monossido di carbonio adsorbito sulla sua superficie; il catalizzatore si raffredda e la reazione ha di nuovo inizio.

La macchina mescolatrice si è dimostrata in grado di riprodurre non solo questa reazione ma anche, e molto bene, quella di Belousov-Žabotinsky. Questa reazione fu scoperta da B. P. Belousov nel 1958 e colpì subito i ricercatori perché, a differenza della maggior parte delle reazioni comuni, non ha uno stato stazionario stabile: oscilla con la precisione di un orologio diventando prima gialla, poi incolore, poi di nuovo gialla e così via, cambiando colore due volte al minuto. I primi a interessarsi della reazione sono stati A. M. Žabotinsky e i suoi collaboratori all'Istituto di fisica biologica presso Mosca; nel 1970 Žabotinsky e A. N. Zaikin hanno riferito di aver osservato onde chimiche circolari che si propagano attraverso una versione modificata del reagente oscillante. Queste onde di colore blu intenso emergono periodicamente formando anelli concentrici simili

TIPO 1 TIPO 2

TIPO 3

TIPO 4

La macchina mescolatrice produce caratteristiche configurazioni a onda

I quattro comportamenti della macchina mescolatrice

a un bersaglio, intorno a punti isolati che fungono da *pacemaker*, in una capsula poco profonda contenente un liquido rosso-arancione.

Nella forma tipica della reazione di Belousov-Žabotinsky, l'acido malonico viene ossidato da bromato di potassio in presenza di un catalizzatore come cerio o ferro. Nella simulazione di Gerhard e Schuster, le cellule della griglia rappresentano, in sostanza, le particelle del catalizzatore, e la metafora del contagio esprime la graduale saturazione della superficie delle particelle.

Ma l'analogia non è così semplice; ci sono alcune sottigliezze. Per dirne una, nella macchina mescolatrice le cellule adiacenti interagiscono passandosi, per così dire, il contagio. In che modo le particelle del catalizzatore scambiano reattività? Secondo Gerhardt e Schuster, almeno nel caso dell'ossidazione del monossido di carbonio, le unità del catalizzatore che prendono parte alla reazione influenzano i loro adiacenti attraverso due meccanismi fondamentali. Una certa unità potrebbe essere resa più reattiva dal trasferimento di calore da un'unità adiacente più attiva o dalla diffusione di monossido di carbonio da un'unità adiacente meno attiva.

L'interazione tra cellule adiacenti nella macchina mescolatrice rende possibile la sincronizzazione della loro attività. Dopo un periodo di iniziale disorganizzazione casuale (la fase di «miscuglio»), le configurazioni che appaiono riflettono questa sincronizzazione. La stessa cosa accade presumibilmente anche nelle vere reazioni chimiche. Può essere, allora, che la macchina mescolatrice spieghi la comparsa di onde di eccitazione nelle reazioni che simula?

Ci sarà chi è pronto a esclamare «Naturalmente!», esibendo come prova le figure. Ma allora, ci sono persone che vedono un automa cellulare in qualsiasi cosa. Sulla rivista «The Atlantic» dell'aprile 1988 c'era un articolo sulle fluttuazioni cosmiche di Edward Fredkin. Fredkin, uomo d'affari nel campo dei calcolatori e ogni tanto professore universitario, suppone che il nostro universo sia composto di cellule che passano da uno stato a un altro come un enorme automa cellulare. A voler essere gentili, si può dire che le prove di questa organizzazione non sono schiaccianti. La macchina mescolatrice è senza dubbio significativa, ma ancor più lo è l'atteggiamento dei suoi inventori. Benché essa simuli in modo notevolmente efficace la reazione di Belousov-Žabotinsky, Gerhardt e Schuster non sostengono che la chimica sia cellulare. Vedono piuttosto il loro automa come uno strumento di approssimazione, la versione discreta di un'equazione differenziale parziale.

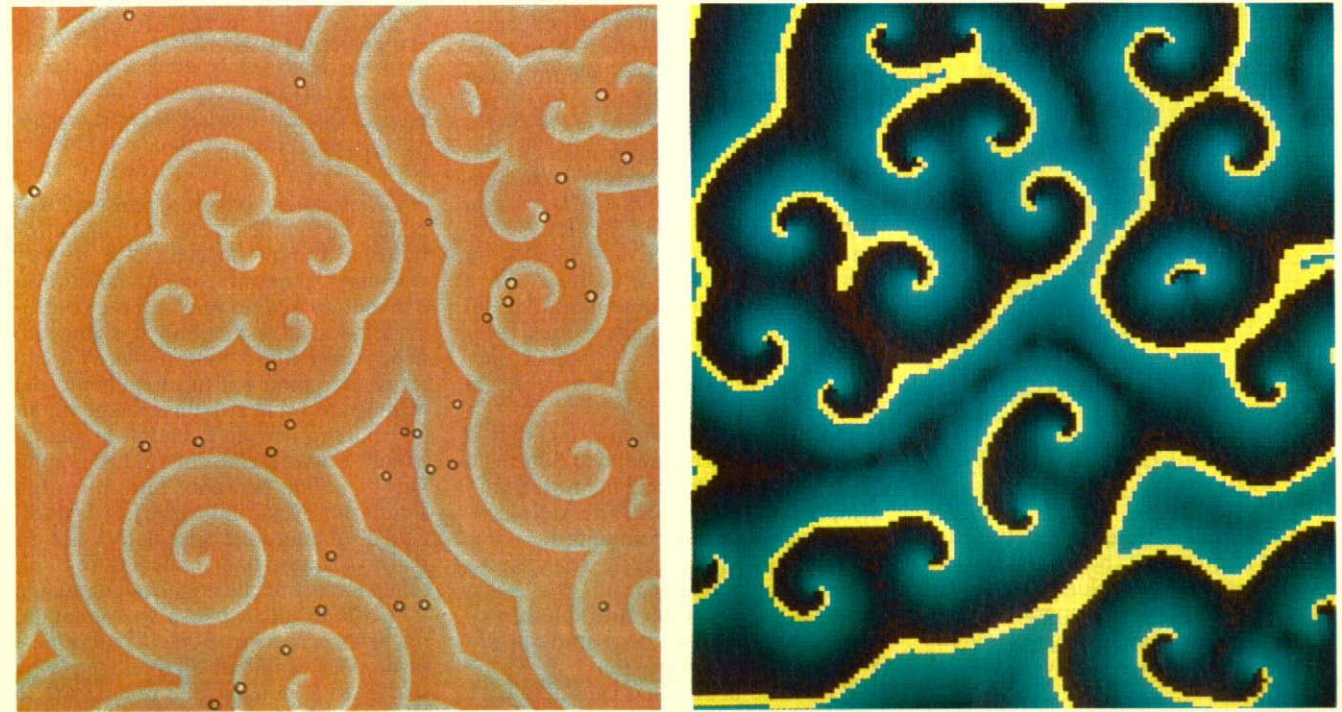
Originariamente ispirati dal lavoro dei chimici Nils Jaeger e Peter Plath dell'Università di Brema, Gerhardt e Schuster ringraziano per l'aiuto nello studio della macchina mescolatrice il loro maestro a Bielefeld, Andreas W. M. Dress, e due chimici, S. C. Müller, dell'Istituto Max Planck per la fisiologia nutrizionale di Dortmund, e John J. Tyson, del Virginia Polytechnic Institute e della Virginia State University. I creatori della macchina vogliono dimostrare che una schiera di oscillatori chimici che interagiscono localmente secondo certe semplici regole finiscono inevitabilmente con il generare onde. È presumibile che ci sia solo un piccolo numero di possibili forme ondulatorie, anche se, a quanto sostiene Tyson, esse diventano molto più complesse nelle tre dimensioni. Dato che i fronti delle onde tridimensionali sono molto più difficili da vedere con gli strumenti di laboratorio, le simulazioni al calcolatore possono dire ai chimici che cosa cercare. Nella scienza, il trucco sta nell'usare bene i modelli, non nell'esserne usati.

I lettori che vogliano costruire da sé la propria macchina mescolatrice hanno già ricevuto ampi suggerimenti sul modo di procedere. Bisogna dichiarare una matrice di dimensione appropriata e incorporarla in un grande ciclo che la aggiorna secondo le tre regole e poi la visualizza per l'edificazione dei «mescolatori». Ciascun elemento della matrice deve contenere il numero di stato per una particolare cellula. Nel calcolare la matrice aggiornata, però, è necessario conservare temporaneamente i risultati in un'altra matrice, fino al completamento del calcolo. A quel punto, un semplice doppio ciclo consente la sostituzione completa della matrice originale con quella aggiornata.

Anche per l'aggiornamento si utilizza un doppio ciclo. Due variabili indice, diciamo *i* e *j*, tengono il conto delle cellule della griglia. Per ciascuna cellula data dalle coordinate (*i*, *j*), il programma (che altro nome dargli se non MISCUGLIO?) decide, tramite una coppia di enunciati condizionali, se la cellula è sana o contagiata. Se è sana, viene valutata la prima formula; se è contagiata, viene valutata la seconda. In entrambi i casi, bisogna controllare lo stato delle cellule adiacenti. Se la cellula non è né sana né contagiata, è ovviamente malata e guarirà al ciclo successivo.

Per ragioni di spazio mi devo limitare a queste brevi indicazioni. I lettori che vogliano una più completa descrizione algoritmica della macchina mescolatrice mi possono scrivere (da qualsiasi paese) presso Scientific American Inc., 415 Madison Avenue, New York, 10017, USA. Si prega di accludere un assegno o un vaglia di 2 dollari per spese postali (in tutto il mondo), fotocopiatura e altri oneri.

L'escursione ad Apraphul ha tratto in inganno poche persone. Quelli che sono comunque entrati nello spirito del resoconto si sono sentiti stimolati dal compito di ricostruire la macchina analogica per moltiplicare di Apraphul: un congegno che moltiplica due numeri usando solo funi e carrucole. Alcuni sono entrati a tal punto nello spirito dell'impresa da sostenere di avere informazioni di prima mano sulla antica cultura di Apraphul. Campione di questa categoria è la lettera inviatami da Clive J. Grant di Chichester, New Hampshire. Un lungo documento descrive la corrispondenza tra Grant e un misterioso dottor Ebur Grebdlog, noto studioso delle antichità di Apraphul:

«Dopo aver letto le sue *(Ri)creazioni al calcolatore* sullo stato della matematica di Apraphul, ho contattato il dottor Grebdlog per chiedergli se nel suo lavoro avesse mai affrontato... un Moltiplicatore Analogico di Apraphul... Mi ha risposto che in effetti aveva fatto degli studi sull'argomento e mi ha inviato una copia del suo lavoro.»

Il «lavoro», scritto con la bella calligrafia di Grebdlog, era accompagnato da disegni tecnici di carrucole e alberi a camme connessi da ponti. Grebdlog annota che i disegni «hanno chiaramente guidato gli abitanti di Apraphul nella costruzione di un congegno veramente notevole per l'inesauribile sforzo tecnologico applicato al suo sviluppo ma ancor più notevole per la sua totale mancanza di utilità».

Il moltiplicatore proposto dal maggior numero di lettori faceva uso di una barra con un'estremità attaccata a un perno fisso. Una fune in ingresso collegata a un punto lungo la barra la tira in avanti in modo che una fune in uscita collegata all'estremità libera viene anch'essa tirata nella stessa direzione. Dato che la barra è in sostanza una leva con il fulcro nell'estremità imperniata, la fune in uscita compie una distanza maggiore della fune in ingresso. Il problema, in questo progetto, è la mancanza di proporzionalità: se si tira ulteriormente la fune in ingresso, la barra segue un arco di cerchio e l'effetto di amplificazione sulla fune in uscita finisce con lo scomparire. Delle variazioni sul tema correggevano a volte il moto circolare della barra attraverso guide o elaborati sistemi di barre parallele congiunte. Tutto ciò mi appariva troppo complicato. Forse avrei dovuto esplicitamente proibire l'uso di barre.

Robert Norton di Madison, Wisconsin, ha usato carrucole a spirale per calcolare logaritmi e antilogaritmi. Le funi in ingresso *A* e *B* sono collegate, srotolate, a due cilindri (che non sono contro le regole, in quanto sono semplicemente delle ampie carrucole). Ciascun cilindro è attaccato a un cilindro a spirale che arrotola una fune in uscita. Le due uscite vengono poi sommate nel modo indicato alla fine dell'articolo di giugno. L'antilogaritmo della somma viene calcolato arrotolando la fune di addizione su un cilindro a spirale collegato a un cilindro diritto su cui è avvolta la fune in uscita finale. Una macchina analoga è stata «scoperta» da Robert A. Eddius di New York City. Gli abitanti di Apraphul, egli sostiene, usavano le conchiglie di certi molluschi con una forma a spirale che li metteva in grado di calcolare esattamente i logaritmi! David A. Fox di Lima, Ohio, ci scrive invece che una cultura simile abitava una piccola isola non lontana dalle Marshall, nota come Hardly Atoll. Qui sono stati trovati non solo gli stessi congegni per il calcolo dei logaritmi e degli antilogaritmi, ma anche un dispositivo simile a uno yo-yo capace di elevare al quadrato i numeri. Forse i lettori vorranno meditare se ciò che Fox afferma sia possibile.

Caxton C. Foster di East Orleans, Massachusetts, è dell'opinione che la civiltà di Apraphul sia stata distrutta dal guadagno logico: il problema incontrato da un calcolatore in cui l'«1» in uscita da ciascuna porta non è proprio un 1. Per impedire a queste imprecisioni di strisciare nei sacri calcoli, i grandi sacerdoti mettevano un abitante di Apraphul a ogni porta per tirare un po' di più le funi in uscita che non avessero la tensione necessaria. Assorbita da questo impegno, la popolazione, non più in grado di procurarsi il cibo, si estinse per inedia.

La parola finale spetta a Michael Pagan di Mount Laurel, New Jersey, un progettista di calcolatori. Preoccupato per il divario culturale tra la branca analogica e quella digitale della società di Apraphul, Pagan ha elaborato un meraviglioso convertitore analogico-digitale. Un'unica fune che porta il segnale analogico entra nel congegno, dal quale esce un gran numero di funi che portano l'equivalente digitale del numero in ingresso. Una macchina di questo genere avrebbe potuto essere introdotta su Apraphul, ma i sacerdoti avrebbero senza alcun dubbio messo al bando questo congegno pagano.

Onde nella reazione chimica di Belousov-Žabotinsky **(a sinistra)** *e loro corrispondenti nella macchina mescolatrice* **(a destra)**

BIBLIOGRAFIA

DEWDNEY A. K., *The Armchair Universe*, W. H. Freeman and Company, 1988.

WRIGHT ROBERT, *Did the Universe Just Happen?* in «The Atlantic», 261, n. 4, aprile 1988.